

Building Platforms to Genetically Encode New Chemistry

Author: Alexander M. Johnson

Persistent link: <http://hdl.handle.net/2345/bc-ir:107292>

This work is posted on [eScholarship@BC](#),
Boston College University Libraries.

Boston College Electronic Thesis or Dissertation, 2017

Copyright is held by the author, with all rights reserved, unless otherwise noted.

Building Platforms to Genetically Encode New Chemistry

by

Alexander M. Johnson

A thesis

submitted to the Faculty of

the department of Chemistry

in partial fulfillment

of the requirements for the degree of

Master of Science

January 2017

Boston College
Morrissey College of Arts and Sciences
Graduate School

Abstract

Unnatural amino acid (UAA) incorporation is a powerful tool used by biochemists to discover the nature of protein structure and function. The evolution of orthogonal aminoacyl-tRNA synthetase (aaRS)/tRNA pairs enables site-specific incorporation of UAAs proteins inside of living cells. The goal of this study was to further expand the repertoire of genetically encoded unnatural amino acids in *E. coli* as well as eukaryotes. We first attempted to engineer an aaRS, previously evolved for *p*-borono-phenylalanine (pBoF), to specifically charge 3-acetyl-*p*-borono-phenylalanine (AcpBoF). A randomized library of the pBoF-specific synthetases was generated and it was subjected to established selection schemes in a bacterial host. This report also describes the development of a yeast-based selection system to alter the substrate specificity of bacterial leucyl-tRNA synthetase, for genetic code expansion in eukaryotes.

Acknowledgements

I would like to thank my advisor Dr. Abhishek Chatterjee for his guidance and patience when dealing with my efforts on my various projects in his lab. I definitely appreciate his kindness and temperance when addressing my works and concerns.

I would like to thank my fellow lab members for their help and encouragement. You all provided just what I needed. I would like to give a large thanks to Ryan for doing the most original experiments regarding the bacteria versus yeast competition and to Anupam for synthesizing the AcpBoF that I used during the experiments.

List of Abbreviations

UAA	unnatural amino acid
pBoF	para-borono-phenylalanine
AcpBoF	3-acetyl- <i>p</i> -borono-phenylalanine
aaRS	aminoacyl tRNA synthetase
Cm	chloramphenicol
Kan	kanamycin
Tet	tetracycline
LRS	leucyl tRNA synthetase
SD	Synthetic Dropout medium
3-AT	3-amino-1,2,4-triazole
5-FOA	5-fluoroorotic acid
noRS	no aminoacyl tRNA synthetase
DanAla	dansylalanine
CAP	2-amino-caprylic acid
AcK	N-acetyl-lysine

Table of Contents

Chapter 1. Introduction to Unnatural Amino Acid Incorporation Systems	1
Research Objectives	4
The Unnatural Amino Acids	5
References	7
Chapter 2. Evolution of pBoF Synthetase to Accept New Residue	8
Abstract	9
Introduction	10
Results	12
Discussion	19
Materials & Methods	20
References	26
Chapter 3. Testing the Yeast Selection System Based on GAL4	27
Abstract	28
Introduction	29
Results	32
Discussion	42
Materials & Methods	43
References	50
Chapter 4. Future Directions	51
Appendix 1. Supporting Information	54
Pictures of Bacterial Selection Plates	55
Selected Yeast Nucleotides	62
Yeast Reporter Information	66
Other Yeast Selection Tests and Spot Plates	69

List of Figures

Fig. 1.1	Structures of pBoF and AcpBoF
Fig. 1.2	Structures of Main UAAs Used in Yeast Selection Tests
Fig. 2.1	Test Expression of pBoF
Fig. 2.2	Preliminary Chloramphenicol Gradient
Fig. 2.3	Spot Plating of pBoF Mock Selection
Fig. 2.4	Spot Plating of AcpBoF Candidates
Fig. 2.5	Spot Plating of pBoF and AcpBoF Candidates from Last Library
Fig. 3.1	Spot Test of DanRS by 3-AT Gradient
Fig. 3.2	Spot Test of DanRS by 5-FOA Gradient
Fig. 3.3	Spot Test of DanRS by Double Positive Selection
Fig. 3.4	Fluorescence Measurements of AcK Gradient
Fig. 3.5	Comparison of Reporter System with or without Additional tRNAs
Fig. 3.6	Proof-of-Concept Positive Selection with AK1 and noRS Constructs with Best tRNA
Fig. 3.7	Proof-of-Concept Negative Selection with AK1 and noRS Constructs with Best tRNA

Chapter 1: Introduction to Unnatural Amino Acid Incorporation Systems

Unnatural amino acids (UAAs) are amino acids that are not canonically used for the ribosomal synthesis of proteins. The ability to site-specifically incorporate UAAs into proteins enable many powerful applications, such as site-specific protein labeling using bioorthogonal conjugation reactions, photo crosslinking to identify elusive protein-protein interactions, incorporation of biophysical and biochemical probe (e.g., a fluorophore), regulation of protein function in living cells, and modeling natural post-translational modifications (1). The incorporation of unnatural amino acids can also be used to evolve new function such as improved enzyme activity (2). Over the last two decades a large number of UAAs have been genetically encoded into bacteria, yeast, and mammalian cells (3).

UAA incorporation can be residue-specific or site-specific. Residue-specific UAA incorporation uses the intrinsic promiscuity of an endogenous aminoacyl-tRNA synthetase towards an UAA, by using a medium lacking the corresponding canonical amino acid (3). However, residue specific UAA incorporation has two major limitations. Since most amino acids are present at multiple sites of most proteins, this approach does not allow precise site-specific protein modification. Additionally, due to its dependence on the promiscuity of the endogenous aaRSs, only close structural analogs can be incorporated, and the efficiency is often very poor (4). In contrast, site-specific UAA incorporation uses an engineered aminoacyl-tRNA synthetase (aaRS)/tRNA pair to specifically deliver the UAA of interest. The tRNA is responsible for recognizing a blank codon and being “charged” (or aminoacylated) with the UAA by its cognate aaRS. Stop codons or quadruplet codons have been used to designate the site of UAA incorporation. TAG is the most commonly used codon for this purpose because of its least use as a stop codon and its high suppression efficiency. One advantage of this technology is that once the engineered aaRS/tRNA pair is available the corresponding UAA can be incorporated in living cells simply by introducing a stop codon in the target protein, and co-expressing the engineered pair in this cell with the mutant target. This technology has been adapted for a variety of organisms and has been optimized for high efficiency, particularly in *E. coli* (3, 5, 6).

This aaRS/tRNA pair must be orthogonal to the other operations inside of the cell (3, 6). This means that the engineered aaRS cannot charge endogenous tRNAs, and that endogenous aaRS cannot charge the engineered tRNA. These aaRS/tRNA pairs are typically imported into a host cell from an evolutionarily distant species to ensure minimal cross-reactivity with their endogenous counterparts. For

example, following aaRS/tRNA pairs have been adapted from various archaeobacteria for use in *E. coli*: tyrosyl, leucyl, glutamyl, lysyl, and prolyl (3). The tryptophanyl and phenylalanyl tRNA/aaRS pairs from *S. cerevisiae* have also been successfully used in *E. coli* (3). Conversely, only the tyrosyl and leucyl tRNA/aaRS pairs from *E. coli* have been adapted for use in eukaryotic hosts such as yeast, and mammalian cells (3). Due to its unique architecture, the archaeobacteria-derived pyrrolysyl tRNA/aaRS pair is orthogonal in both prokaryotic and eukaryotic cells, providing a versatile platform for genetic code expansion (3, 7).

Once an efficient, orthogonal tRNA/aaRS pair is available to suppress a nonsense/frameshift codon, its native amino acid specificity can be altered by engineering the aaRS active site. Structural information is used to generate a large library of aaRS active site variants by randomizing amino acid residues that participate in substrate-recognition. To isolate the unique variants capable of charging a ncAA of interest, but none of the canonical amino acids, this library is subjected to a double-sieve selection scheme (3). The first round of selection directly couples the activity of the aaRS to the survival of the expression host, leading to the enrichment of active aaRS variants capable of charging either the desired ncAA, or one of the canonical amino acids. This “positive selection” typically employs an antibiotic or metabolic selection marker harboring an inactivating TAG stop codon at a permissive site, such that the expression of full length protein, and the survival of the host under selective conditions, is contingent upon the presence of an active TAG-suppressing tRNA/aaRS pair. To eliminate the aaRS variants charging canonical amino acids from this enriched pool, a second selection step is used that couples the activity of the aaRS variant to the expression of a toxic protein leading to host cell death. The toxic gene is inactivated with one or more TAG codons at permissive sites, and is only expressed in the presence of an active TAG-suppressing tRNA/aaRS pair. The aaRS variants capable of charging canonical amino acids would facilitate the expression of this toxic gene, and consequent cell death, if this “negative” selection is performed in the absence of the desired ncAA, while the ncAA-specific variants would be unable to express the toxic gene under these conditions. This strategy has been successfully utilized to genetically encode a large number of amino acids in both prokaryotes and eukaryotes. *E. coli* and *Saccharomyces cerevisiae* (yeast) are used as hosts for selection of aaRS variants to be used in prokaryotes or eukaryotes, respectively. In *E. coli*, an antibiotic resistance gene, typically chloramphenicol acetyltransferase (CAT), is used for positive selection,

while the toxic RNase barnase is used as the negative selection marker. In yeast, the metabolic marker HIS3 is used for positive selection, while URA3, which leads to the conversion of 5-fluoroorotic acid to toxic 5-fluorouracil, has been used as the negative selection marker. Typically, engineered suppressor tRNA/aaRS pairs developed in *E. coli* or yeast can be used in other prokaryotic or eukaryotic organisms, respectively, by expressing these from appropriate promoters. For example, tRNA/aaRS pairs evolved in yeast are directly imported into various mammalian cells, where similar selection experiments are challenging to execute due to various technical limitations.

The success of UAA incorporation also depends on how well the orthogonal aaRS/tRNA uses the native translational machinery, such as elongation factors, release factors, and the ribosome. EF-Tu is the protein that is responsible for shuttling aminoacylated tRNAs to the ribosome and the acceptance by EF-Tu improves the outcome of UAA incorporation (8). At the stop codon sequence, release factors (RFs) bind to the mRNA to facilitate the release of the ribosome and the removal of RF1 (whose exclusion specifically affects TAG in this case) removes competition for the stop codon sites to promote UAA incorporation (9). Engineering these factors can lead to further improvement in UAA incorporation efficiency (3).

Although many UAAs have been genetically encoded so far, a variety of chemical functionalities still remain elusive. One particularly useful chemical functionality is that encoded by 3-acetyl-*p*-boronophenylalanine (Figure 1.1), which enable reversible covalent attachment to lysine residues under physiologically conditions (10). Previously, para-borono-phenylalanine (pBoF) unnatural amino acid was genetically encoded using the *M. jannaschii* tyrosyl tRNA/aaRS pair in *E. coli* (11). One of the goals of the current study is to start with this evolved tRNA/aaRS pair to evolve a variant specific for 3-acetyl-para-borono-phenylalanine (AcpBoF). The ability to genetically encode this unique amino acid may enable the evolution of new protein functions that takes advantage of its ability to form reversible-covalent association with lysine.

Finally, this project also focuses on the development of a yeast selection system to engineer the leucyl-tRNA synthetase/tRNA pair from bacteria to genetically encode new UAAs in eukaryotes. The leucyl platform was previously used to genetically encode a small but structurally diverse set of novel UAAs (Figure 1.2), many of which have been known to operate efficiently at low concentrations of supplementation (3). However, no new UAAs have been genetically encoded using this platform in last

seven years. Further development of this platform will provide new genetically encoded UAAs that remain inaccessible to other aaRS/tRNA pairs. This platform may also be combined with the archaeal pyrrolysyl platform to site-specifically incorporate two distinct UAAs into proteins in mammalian cells.

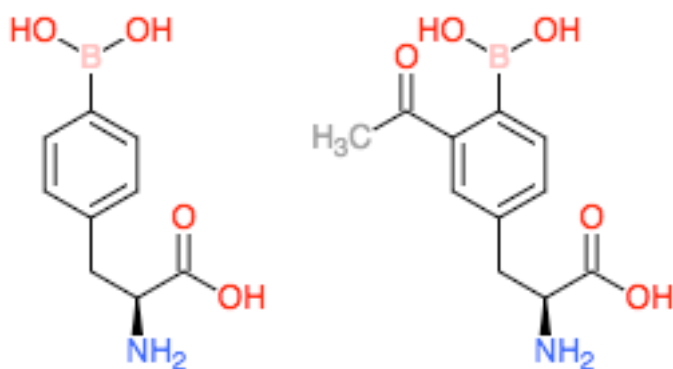


Fig. 1.1 Structure of pBoF and AcpBoF. The pBoF unnatural amino acid was already incorporated in *E. coli* with a YRS/tRNA pair evolved from *M. jannaschii*. The AcpBoF (right) unnatural amino acid is a potentially new member of incorporation.

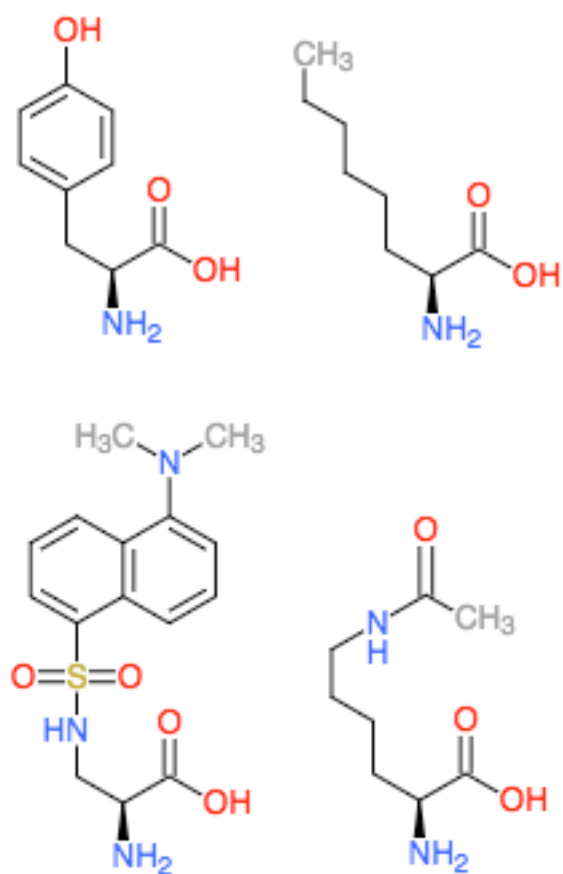


Fig. 1.2 Structures of some of the UAAs genetically encoded using the leucyl platform. Top left: O-methyl-tyrosine. Top right: 2-amino-caprylic acid. Bottom left: dansylalanine. Bottom right: N-acetyl-lysine.

References

- (1) Lin, S.; Zhang, Z.; Xu, H.; Li, L.; Chen, S.; Li, J.; Hao, Z.; Chen, P. R. *J. Am. Chem. Soc.* **2011**, *133*, 20581-20587
- (2) Ugwumba, I. N.; Ozawa, K.; Xu, Z.; Ely, F.; Foo, J.; Herlt, A. J.; Coppin, C.; Brown, S.; Taylor, M. C.; Ollis, D. L.; Mander, L. N.; Schenk, G.; Dixon, N. E.; Otting, G.; Oakeshott, J. G.; Jackson, C. J. *J. Am. Chem. Soc.* **2011**, *133*, 326-333
- (3) Liu, C. C.; Schultz, P. G. *Annu. Rev. Biochem.* **2010**, *79*, 413-444
- (4) Singh-Blom, A.; Hughes, R. A.; Ellington, A. D. *J. Biotechnol.* **2014**, *178*, 12-22
- (5) Chatterjee, A.; Sun, S. B.; Furman, J. L.; Xiao, H.; Schultz, P. G. *Biochem.* **2013**, *52*, 22071-22083
- (6) Chin, J. W. *Annu. Rev. Biochem.* **2014**, *83*, 379-408
- (7) Fekner, T.; Chan, M. K. *Curr. Opin. Chem. Biol.* **2011**, *15*, 387-391
- (8) Jeong, K.; Pavlov, M. Y.; Kwiatkowski, M.; Ehrenberg, M.; Forster, A. *RNA* **2014**, *20*, 632-643
- (9) Johnson, D. B.F.; Xu, J.; Shen, Z.; Takimoto, J. K.; Schultz, M. D.; Schmitz, R. J.; Ecker, J. R.; Briggs, S. P.; Wang, L. *Nat Chem Biol.* **2012**, *7*, 779-786
- (10) Bandyopadhyay, A.; McCarthy, K.A.; Kelly, M.A.; Gao J. *Nat Commun.* **2015**, *6*, 6561
- (11) Ryu, Y.; Schultz, P. G. *Nature Methods* **2006**, *3*, 263-265

Chapter 2: Evolution of pBoF Synthetase to Accept New Residue

Abstract

The *M. jannaschii* tyrosyl tRNA synthetase/tRNA pair was previously evolved to genetically encode pBoF in *E. coli*. This synthetase is a good starting point for the directed evolution of a variant that can incorporate AcpBoF (Figure 1.1). In this study, the mutagenesis of the pBoF synthetase was used to create libraries, which were subjected to directed evolution experiments. The selection system was first verified by using this library to identify variants that charge pBoF. All selected clones were found to selectively incorporate pBoF in response to a TAG codon. However, a number of attempts to use these libraries to identify AcpBoF variant did not succeed, suggesting that the pBoF synthetase may not be a suitable starting point for developing our desired variant. A much larger and completely naïve active site library should be used to develop an engineered aaRS specific for this UAA.

Introduction

The UAA para-borono-phenylalanine (pBoF) has previously been incorporated in bacterial cells using a *M. jannaschii* tyrosyl Trna synthetase/tRNA pair (1). The evolved pair was demonstrated to successfully incorporate pBoF into proteins and enable several downstream applications. pBoF is known for its abilities to bind to diols, sugars, spatially close cluster of serine residues, etc. pBoF can be used in proteins to evolve novel binding domains with specificity for glycans (2). This chapter explores the effort of trying to evolve the pBoF-specific tRNA/aaRS pair to be able to incorporate another UAA, 3-acetyl-p-borono-phenylalanine (AcpBoF). AcpBoF is a modified version of pBoF and has the capability to form strong bonds with amines via boronic coordination of the amine to the acetyl group (3).

Although no crystal structure of the pBoF synthetase is available, we used the structure of the wild-type *M. jannaschii* tyrosyl synthetase (4) as a guide to design the mutant library. In order to free space in the pBoF synthetase active site to accommodate the new acetyl group in AcpBoF, the L69, M70, and Q109 residues were selected as targets for randomization. The overall diversity of a mutant library depends on how many residues are mutated and to which codons. This study used NNK codons to randomize each residue, and each NNK codon contributes 32 different variants to the diversity of the overall library. The NNK codon (N = nucleotide = A or G or T or C)(K = G or T) does not contain every potential nucleotide combination. However, the NNK codon still covers all possible types of residues and can be used in place of the NNN codon, whose contribution to diversity is 64 . In the case of the BoF library, the L69 and M70 residues are adjacent and are easily mutated by a single mutagenesis reaction. This library, termed the 1-site BoF library, has a diversity of $32 \times 32 = 1,024$. The 2-site BoF library, which also includes Q109, has a diversity of $32 \times 32 \times 32 = 32,768$. It is important to generate the library with an efficiency such that all possible members are represented.

The plasmids used in this bacterial selection system are pBK, pRep, and pNeg, which are the plasmids that contain the synthetase of interest, the positive selection machinery, and the negative selection machinery, respectively. The library members are encoded in the pBK plasmids that are transformed into cells harboring either pRep or pNeg plasmids, for positive and begative selections, respectively. These selection plasmids also encode the cognate tRNA for the aaRS being engineered. This plasmid system have been previously engineered for efficient selection of aaRS/tRNA pairs in *E. coli* (5).

The pRep plasmid, encodes a chloramphenicol resistance gene with 1 TAG at a permissive site. When an amber-suppressing aaRS/tRNA pair is expressed along with their substrate UAA, the TAG sites are filled and Cm resistance is expressed. Therefore, active aaRS mutants will be able to grow on chloramphenicol with proportion to their activities. The pRep plasmid also contains a GFPuv under a T7 promoter, as well as a T7-RNA polymerase that has two TAG mutations (6). Thus, in the presence of an active amber suppressing aaRS/tRNA pair, the T7-RNA polymerase gene is expressed and drives strong overexpression of GFPuv. Active mutants will appear green in response to UV light, providing a secondary verification for the activity of the aaRS/tRNA pair.

The pNeg plasmid also encodes the cognate tRNA and also expresses a barnase gene, the expression of which is toxic to the cells. The barnase gene is inactivated with TAG codons at permissive sites, and in the presence of an active aaRS/tRNA pair it is expressed and leads to cell death. The negative selection is performed in the absence of the target UAA, such that any aaRS mutant that charges a canonical amino acid will be eliminated due to barnase expression, while the aaRS mutants selective for the UAA would not. In this case, since we chose to use the pBoF synthetase as a starting point, which is already orthogonal due to previous evolutions, the negative selection was found to be unnecessary.

Results

Validating the pBoF specific aaRS

Before attempting to perform the selection with AcpBoF, we wanted to ensure that the selection components are working. To first verify the activity of the pBoF synthetase, it was expressed from a pEVOL vector along with its cognate tRNA, and a super-folder GFP (sfGFP-151-TAG) reporter protein was coexpressed from a pET22b plasmid. The test expression showed robust expression of the sfGFP reporter only in the presence of the pBoF but not its absence (Figure 2.1). This result confirms that the aaRS that we constructed specifically charges pBoF in response to TAG stop codons, and that it charges other UAAs at a very low level.

Next, we sought to determine the chloramphenicol concentration that produce a clean difference of bacterial growth with or without the UAA during the positive selection. Without pBoF, the cells expressing the pBoF-specific aaRS grew up to 15 µg/mL chloramphenicol (Cm), while in the presence of 1 mM pBoF the cells grew up to 50 µg/mL (Figure 2.2). A solid difference based on UAA is provided at the 50 µg/mL well, but valid concentrations to use in selection can also exist above 15 µg/mL chloramphenicol. Thus, the range of 15-50 µg/mL chloramphenicol (Cm) provides a suitable concentration range where selection of active aaRS should be possible.

Mock selection of the aaRS library to isolate pBoF specific variants

The mock selection of 1-site mutant BoF synthetase library was performed using 30 µg/mL chloramphenicol. The plate with no Cm or pBoF (control plate with no selective pressure) demonstrated robust growth. The selection plate (+Cm) with pBoF plate was less densely filled with colonies, but in the absence of pBoF no significant growth was observed (except for a few cheaters). Thus, the mock selection plates were behaving as anticipated. Twenty-four colonies were picked from the +pBoF plate to be used in a spot testing in 50 µg/mL chloramphenicol with or without 1 mM pBoF. All 24 spots grew in the presence of pBoF and did not grow without pBoF (Figure 2.3).

First Selection Attempt for AcpBoF

The 2-site BoF synthetase library was completed before the first selections with AcpBoF. The selection used two different concentrations of chloramphenicol, 30 µg/mL and 50 µg/mL to ensure isolation of even weakly active aaRS variants with the desired activity. The negative control plates (-

AcpBoF) had a very small number of colonies on them, with a greater number being present on the plate with less Cm concentration. The +pBoF control plates were populated with colonies as expected from the mock selection at both concentrations of the antibiotic. The +AcpBoF plates had a very small number of colonies on both plates and these colonies had irregular sizes. All of the colonies on the +AcpBoF plates were subjected to spot testing with 50 µg/mL chloramphenicol and \pm AcpBoF. The spot testing did not reveal differential growth in the presence or the absence of the UAA, and some of the colonies did not grow on chloramphenicol at all (Figure 2.4).

Second Selection Attempt for AcpBoF

The other selections were done with 20 µg/mL and 30 µg/mL chloramphenicol with the new 2-site pBoF synthetase library that was ensured to have complete coverage and was confirmed to be free of bias. As seen before, the +pBoF control plates with this library were appropriately populated with colonies on the selection plates. In this case, the –UAA plates had several unwarranted colonies of at least two distinctively different sizes (which is expected due to the low selection pressure). The +AcpBoF plates are populated in a similar way as the –UAA plates are, which was indicative of unsuccessful selection.

Ten colonies from each +pBoF and +AcpBoF plate were still sampled for further characterization. When the colonies were spot tested against their respective UAAs, 19 of 20 pBoF-selected colonies exhibited differential growth in the presence or absence of pBoF, whereas only 1 of 20 AcpBoF-selected colonies showed any difference at all (Figure 2.5). Most of the AcpBoF colonies grew on chloramphenicol irrespective of the presence of the UAA. However, upon further testing the only promising colony also failed to show successful AcpBoF incorporation.

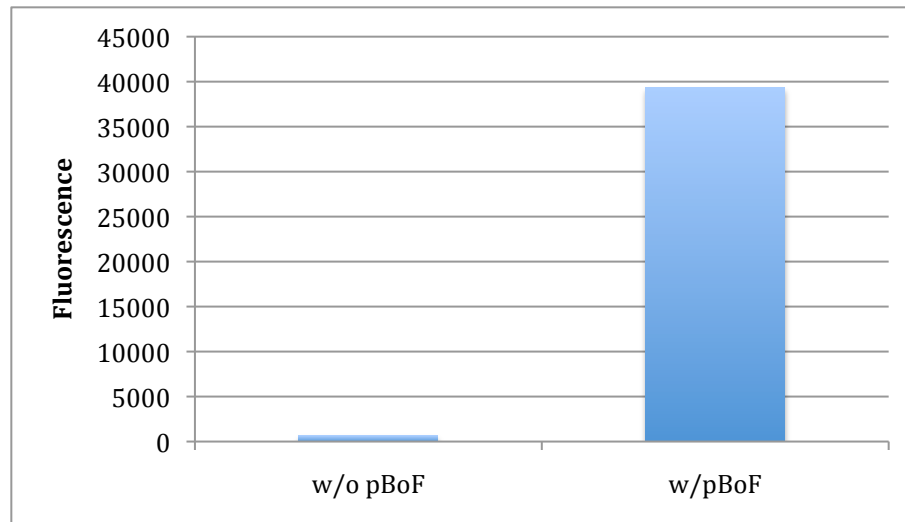


Fig. 2.1 Expression of sfGFP with or without 1 mM pBoF. The fluorescence of *E. coli* cultures transformed with pEVOL-pBoF and pET22b-t5-sfGFP-151-TAG was recorded after protein expression.

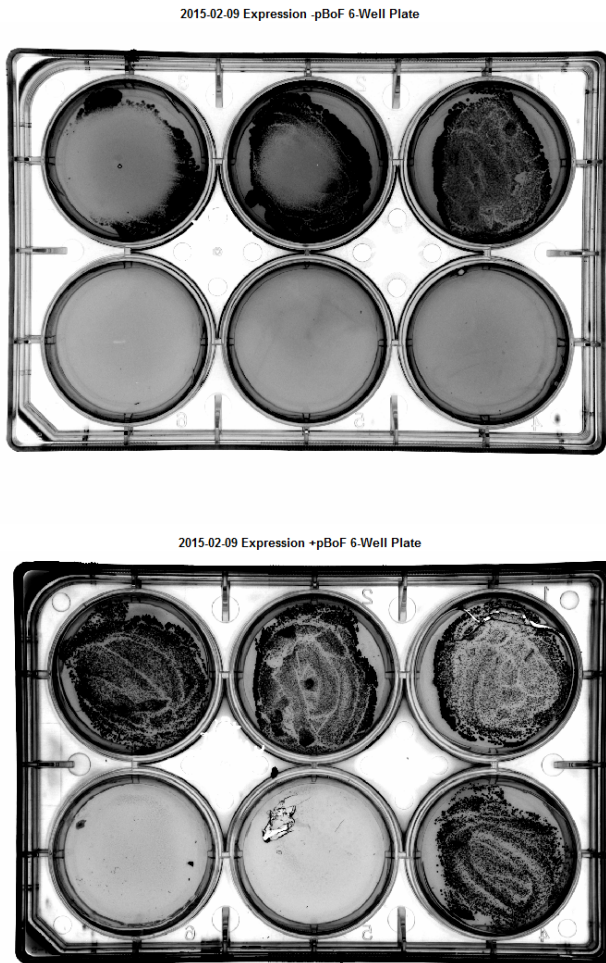


Fig. 2.2 Preliminary chloramphenicol gradients with *E. coli* pBK-pBoF + pRep-MjY-CUA. The top plate corresponds to the -pBoF media and the bottom plate corresponds to the + 1 mM pBoF media. All wells contain LB+KAN+TET+Arabinose agar with different levels of chloramphenicol and were spread with 5 μ L of $OD_{600} = 0.01$ cells. The concentrations of chloramphenicol were analogous for both plates and were enumerated as, from the top right well to the bottom left well horizontally, 0 μ g/mL Cm, 5 μ g/mL Cm, 15 μ g/mL Cm, 50 μ g/mL Cm, 100 μ g/mL Cm, and 200 μ g/mL Cm. The 50 μ g/mL Cm well had a difference of growth based on the presence of the UAA.

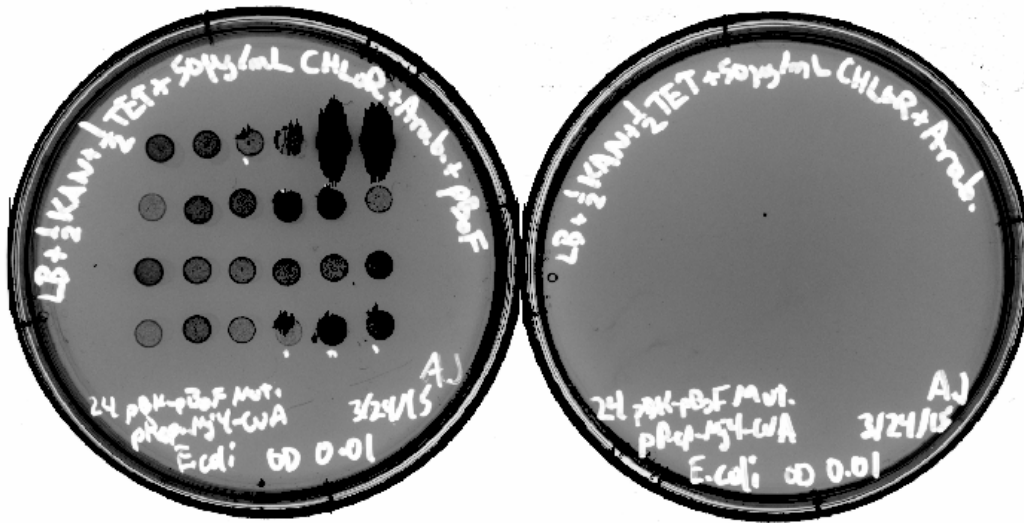


Fig. 2.3 Spot testing of 24 selected mutants from mock selection of 1-site pBoF synthetase library. Each spot represents 5 μ L of $OD_{600} = 0.01$ *E. coli* cells with mutant pBK-pBoF and pRep-MjY-CUA and the spots were made analogously on the two plates, which differ by the presence and absence of 1 mM pBoF. All 24 mutants showed UAA dependent growth at 50 μ g/mL Cm.

2015-04-13 AcpBoF Spot Testing

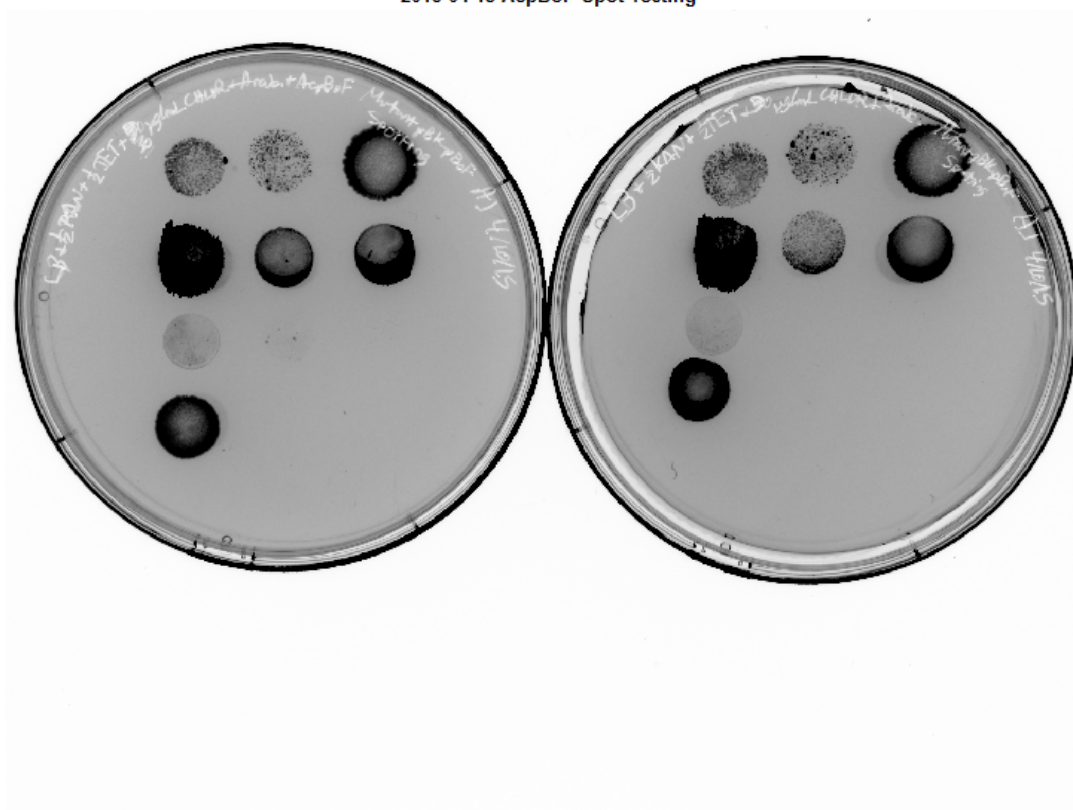


Fig. 2.4 Spot testing of 15 selected candidates from +AcpBoF plates of first-round selections. Each spot represents 3 μ L of $OD_{600} = 0.01$ *E. coli* cells harboring mutant pBK-pBoF and pRep-MjY-CUA. Both plates contain LB + 1/2KAN + 1/2TET + Arabinose + 30 μ g/mL Cm agar; the left plate contains +1mM AcpBoF and the right plate does not contain a UAA. Eight of the colonies grew irrespective of the AcpBoF in the medium and seven colonies did not grow under either condition.

2015-04-15 Spot Testing

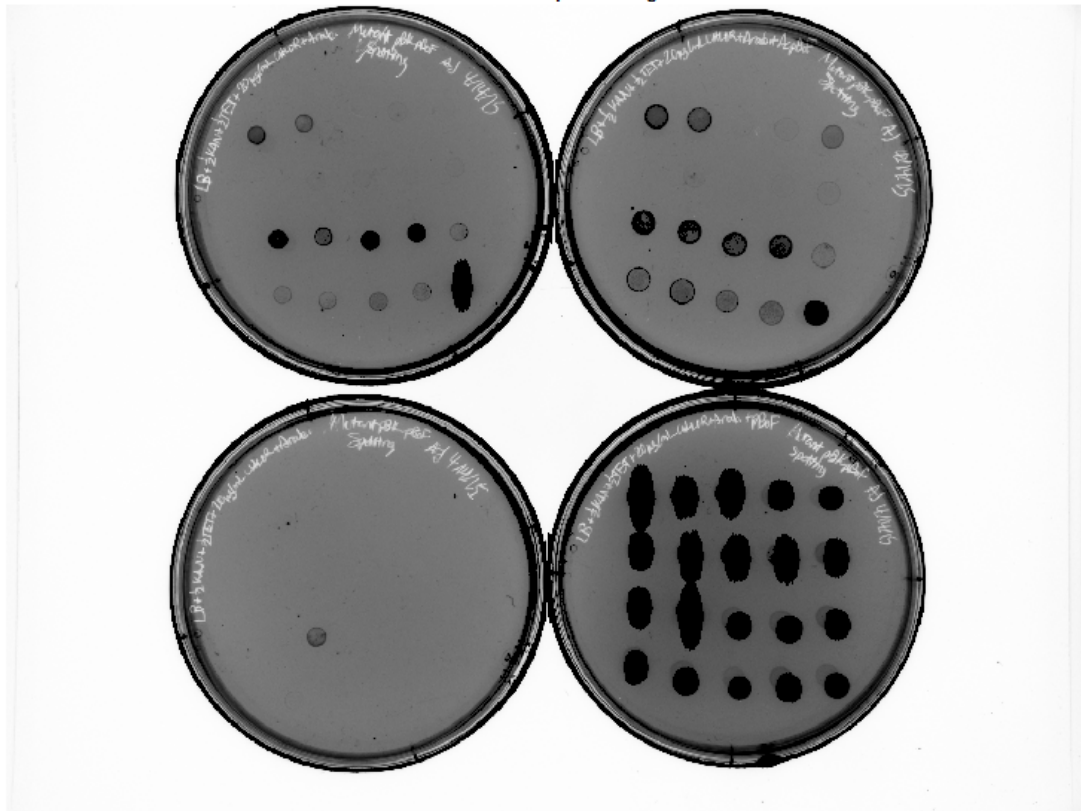


Fig. 2.5 Spot testing plates comparing the UAA differences of +pBoF clones and +AcpBoF clones during the same selections. Bottom: Twenty +pBoF *E. coli* clones (ten from each second-round selection) were plated on -pBoF (left) and + 1 mM pBoF (right) plates in form of 3 μ L OD₆₀₀ = 0.01 spot each. All media contained LB + 1/2KAN + 1/2TET + Arab. +20 μ g/mL Cm. The spots were positioned analogously on the corresponding plates. Top: Analogous plates for the +AcpBoF clones. Almost all +pBoF clones showed a whole difference based on UAA, but the +AcpBoF clones showed weakly discernible to no difference.

Discussion

There Are No Synthetase Mutants That Incorporate AcpBoF

The selections and trials verify that the pBoF synthetase cannot be mutated to incorporate the closely related unnatural amino acid AcpBoF. The pBoF synthetase appeared to be an obvious platform to be used as a starting point for evolving a variant that charges the new UAA. Based on the expression testing and mock selection, we are confident with the coverage of the library as well as the integrity of the selection system. One assumption we made while designing the library was that the aaRS specific for pBoF would be analogous to the wild-type aaRS. Since the structure of the mutant aaRS was not available, we used the wild-type structure to construct our library. It is possible that this assumption was inaccurate and that the pBoF binds in the active site differently than tyrosine binds to the wild-type active site. This would render our library design, trying to make space for the extra acetyl group, incorrect leading to the failure of identifying a desired aaRS.

This does not necessarily verify that developing an AcpBoF specific incorporation is impossible. However, it may be necessary to start from scratch by building a large library of the wild-type aaRS by randomizing key active site residues that could lead to the accommodation of the UAA AcpBoF. Construction and selection of such very large naïve library is challenging but may be necessary to evolve unique aaRS variants capable of charging this unique aaRS.

Materials & Methods

General

Cells were imaged on a ChemiDoc MP system. Chemicals and reagents were commercially purchased from Fisher Scientific or Sigma-Aldrich. PCR primers were purchased from IDT. Sequencing was done at Eton Bioscience. PCR reactions were carried out with Phusion Hot Start II High-Fidelity DNA polymerase, except where *Thermus aquaticus* DNA polymerase is noted. PCR reactions followed the manufacturer protocol except where noted otherwise. Ligation reactions were done with T4 DNA ligase in its respective buffer. T4 DNA ligase and restriction enzymes for cloning were purchased from NEB. Absorbances were taken on Nanodrop 2000c Spectrometer. Fluorescence measurements were taken on SpectraMax M5 plate reader. Electroporations were done on Bio-Rad Gene Pulser.

Test Expression of pBoF System

The pEVOL-pBoF and pET22b-t5-sfGFP-151-TAG plasmids were cotransformed into Top 10 *E. coli* with a chemical transformation protocol. I inoculated a 10 mL LB+AMP+CHLOR culture with 100 μ L of starter culture and grew to OD of 0.612 at 37 °C with shaking. At that time, 100 μ L 100 mM IPTG and 10 μ L 20% arabinose were added as part of induction. The culture was separated into two 5 mL cultures, into 1 of which 50 μ L 100 mM pBoF was added. After overnight growth and expression, 300 μ L samples of each culture were taken as samples and centrifuged at 5,000 x g for 8 minutes. The pellets were resuspended in 200 μ L PBS and then diluted ten-fold in PBS on a 96-well plate. The fluorescence intensities of the wells were read with a plate reader.

Chloramphenicol Screening Gradient

E. coli with pBK-pBoF and pRep-MjY-CUA were inoculated into 5 mL LB+TET+Kan culture and grown overnight at 37 °C. The culture was diluted to an OD of 0.01 before spreading 5 μ L each on 2 6-well plates of LB+TET+KAN+arabinose with chloramphenicol concentrations of 0, 5, 15, 50, 100, and 200 μ g/mL in each of the 6 wells in series. The medium of one plate lacked pBoF and the medium of the other plate contained 1 mM final pBoF. After 2 days of growth, the plates were imaged.

Cloning of 1-Site BoF Library

The BoF synthetase was cloned from the pEVOL-pBoF plasmid using the restriction enzymes NdeI-HF and KpnI-HF to amplify out the synthetase and insert it into the pBK-MCS-MjYRS plasmid. The

synthetase insert and pBK vector were ligated in a reaction with T4 ligase and electroporated. After the plasmid was isolated from the colonies, the pBK-pBoF plasmid was subjected to mutagenesis to create the library. The library that mutated the L69 and M70 sites applied PCR to amplify two different fragments of the synthetase, which was divided into a mutant N-terminal fragment and an overlapping C-terminal fragment. The two fragments were combined in a primerless overlap extension PCR, whose whole product was gel purified. This DNA product was used as the template for further PCRs with terminal primers with respect to the pBoF synthetase. After amplifying the mutant synthetase, the synthetase and pBK-pBoF plasmids were digested with BamHI-HF and NcoI-HF to create the inserts and vectors to use for the library ligation and electroporation. A small-scale ligation reaction and electroporation was performed as a test before doing the large-scale reactions and electroporations used for the library. After efficient electrocompetent cells were made, the series of electroporations were done with a serial dilution plating. The electroporation and processing of bacterial libraries was done as described in following sections.

Cloning of 2-Site BoF Library

The mutagenesis of the Q109 site was performed on the template of the LM-site mutant BoF library DNA. A fragment that is downstream of the previous mutations was amplified by a mutagenic forward primer and terminal reverse primer. The PCR products and mutant pBK-pBoF library plasmid were digested with EcoRI-HF and NcoI-HF in order to generate the inserts and vectors used to replace part of the synthetase with the mutated version. The vectors and inserts were ligated and then the library was processed according to the procedures in other sections.

Recloning of 2-Site BoF Library

The second LMQ-mutated BoF library was created in terms of 3 fragments, 1 of which contains the L69 & M70 mutations (341 bp), 1 of which has the Q109 mutation (151 bp), and 1 of which is the wild-type C-terminal remainder of the synthetase (608 bp). The first fragment is amplified with terminal forward primer and mutagenic reverse primer; the second fragment is amplified with forward primer that overlaps with the downstream flank of the previous reverse primer and another mutagenic reverse primer; the third fragment is amplified with another overlapping forward primer and a terminal reverse primer. Different OE PCR protocols were tested and the final OE PCR protocol used was one where the LM-site and Q-site fragments were reacted for half of the cycles and then the C-terminal fragment was added to the PCR

mixtures for the other half of cycles. The whole library DNA was amplified by Taq polymerase and otherwise the preparation of library was done according to the other sections.

Ethanol Precipitation of Ligated Library DNA

For 1.6 µg of contributed DNA in a 100 µL ligation reaction, the protocol contributes yeast tRNA to increase the amount of DNA that is visible upon centrifugation. The ligation reactions, yeast tRNAs, and ongoing samples are kept on ice unless manipulated otherwise. Each ligation reaction is transferred into a 2-mL Eppendorf tube via pipette. Then 5 µL of 10 mg/mL yeast tRNA and 250 µL of absolute ethanol (200 proof) is added to the reaction (for a final EtOH% of 70%). Each tube is mixed and then incubated at -20 °C for 30 minutes to precipitate DNA. This mixture is centrifuged at 16,000 x g for 30 minutes at 4 °C. The supernatant is carefully and completely discarded from the tube. Then 250 µL of 70% ethanol (made from absolute ethanol 200 proof and ddH₂O) is added to the DNA smear. The DNA is gently dislodged from the tube by pipetting and then centrifuged under the same conditions again. The supernatant is discarded again. The sample tubes are inverted on a Kimwipe for a few minutes and then stand for 1 hour in the cold room with caps open to air dry. The final pellet is resuspended in 3-5 µL of ddH₂O or 10 mM Tris-HCl pH 8.5 and stored on ice or for long term at -80 °C until the electroporation.

Library Electroporations

Both pBoF and LRS libraries were electroporated into *E. coli*, whose strain was either Top 10 or EMAX DH10B. The ethanol-precipitated and resuspended ligation mixtures are mixed with electrocompetent cells on ice. Generally, 1 or 5 50 µL electrocompetent cell samples were used for every ligation reaction based on 1.6 µg of contributed vector and insert DNAs. 1-mm electrocuvettes were briefly chilled on ice before usage. In some cases, the recovery medium (1 mL LB broth) was warmed to 37 °C before usage. For each electroporation, 50 µL of cell-DNA mixture is added to the electrocuvette's slit directly. The electrocuvette is tapped and flicked to ensure that the mixture settles at the bottom of the cuvette and the cuvette is also dried with a Kimwipe before placing the cuvette into the slider. The electroporation takes place at 1.8 kV, 200 Ω. and 25.0 µF. Immediately after pulsing ends, the cuvette is removed from the system and 1 mL of LB broth is pipetted into the cuvette. The pipette is used to gently mix the solution and transfer it to a culture tube. The addition, settling, pulsing, and pipetting process is done as quickly as possible, especially after the pulse, and is done as many times as there are

electroporations. Each 1-2 electroporations at a time is transferred to a 37 °C shaker for recovery at 250 rpm for 1 hour.

After the recovery, library cultures are mixed together thoroughly. A 10 µL sample is pipetted into 990 µL of sterile ddH₂O and another 10 µL sample is pipetted into 90 µL ddH₂O. 100 µL of the 100-fold dilution is pipetted into 900 µL of sterile ddH₂O. This process is repeated two more times across the solutions. On 5 different selective plates (LB+Kan or LB+ampicillin), 100 µL of recovery solutions of the different dilution concentrations are added, to represent 10 µL of the recovery culture as well as 10², 10³, 10⁴, and 10⁵ dilutions from that 10 µL. For the remaining combined recovery, 1 mL of culture is plated onto 1 large selective plate until there is no more volume. All plated volumes are spread onto the plates. The Bunsen burner is used to flame a spreader dipped in ethanol before and after every spreading. The large agar plates are left by the flame until dried. All plates are incubated upside-down at 37 °C.

Characterization of Bacterial Libraries

The library electroporations are grown overnight at 37 °C. The large and small plates are imaged. For the dilution series, the number of colonies on the sparser plates is counted manually as an estimate of the transformation efficiency. If the transformation efficiency is high enough, then the large agar plates are harvested. To harvest, by the Bunsen burner flame, pipette 10 mL of LB broth to each plate and then spread the broth into the colonies. Spread as much as of the plate as possible to remove visible traces of colonies on the agar and to disperse cellular streaking in the medium. Use a different 10 mL pipette to intake the cellular suspension and contribute to 50-mL conical tube. Repeat the process of adding LB broth, spreading, and retrieving cells until all large plates are harvested. Thoroughly mix the 50-mL conical tube with the library of cells. Use 10 µL of the final solution in 990 µL LB broth, against 1 mL LB broth as a blank, to record the 100-fold dilution of the OD₆₀₀ of the culture. Derive the full OD₆₀₀ and calculate how much volume of culture is used for each miniprep of library DNA ($3 \text{ OD}_{600} \times 5 \text{ mL} = _ \text{OD}_{600} \times _ \text{mL}$). A few aliquots of the library culture are saved as glycerol stocks and about 10 volumes were miniprepped for an ordinary library. After miniprepping and analysis, library DNA gets combined and the good plasmids of 10 or more individual colonies are submitted for sequencing to check the diversity of the library.

Bacterial Miniprep Protocol

Pick a single colony from the desired plate and inoculate it with the pipette tip into 5 mL of selective LB medium. Let the culture tubes grow overnight at 37 °C. For library harvests, the designated volume of culture is directly added to 1.5-mL Eppendorf tube and centrifuged as the first step. Centrifuge the culture tubes at 4,500 rpm for 10 minutes at 4 °C. Discard supernatants and add 250 µL of PB1 buffer to each tube. Resuspend the pellets and transfer to 1.5-mL tube. Add 250 µL of PB2 to each tube and shake vigorously for 15-20 seconds. Let the solution stand for 1 minute. Add 300 µL of N3 solution to each tube and shake for 15-20 seconds. Centrifuge those tubes at 16,000 x g for 10 minutes. Transfer the supernatants as carefully as possible to miniprep columns in collection tubes. Centrifuge the tubes at 11,000 x g for 20 seconds. Discard liquid in collection tube and add 500 µL PB4 to each miniprep column. Centrifuge at 11,000 x g for 20 seconds again and discard contents of collection tube. Add 600 µL PB5 to each tube and centrifuge at 11,000 x g for 20 seconds again. Discard contents of collection tube and move miniprep columns into fresh 1.5-mL Eppendorf tubes. Add 50 µL of PB6 buffer to each miniprep column. For this addition, use a new pipette tip for every column and eject liquid directly onto the silica membrane. Let the columns stand for 1 minute. Centrifuge at 11,000 x g for 30 seconds to elute your DNA. Tap the tubes of DNA to mix them together.

Bacterial Miniprep Reagents

PB1 solution is 50 mM tris-HCl at pH 8.0, 10 mM EDTA, and 100 µg/mL RNase A. PB2 solution is 200mM NaOH and 1% SDS. N3 solution is 4.2 M guanidinium hydrochloride plus 0.9 M potassium acetate at pH 4.8. PB4 is 5.0 M guanidinium hydrochloride in 30% isopropanol. PB5 solution is 10 mM tris-HCl at pH 8.5 in 80% ethanol. PB6 solution is 10mM tris-HCl pH 8.5, with or without 1mM EDTA.

Agarose Gel Electrophoresis

To create a 1% (w/v) agarose gel in TAE buffer (40 mM tris, 20 mM acetate, 1 mM EDTA at pH 8.0), the respective mass of agarose and volume of TAE is added to an Erlenmeyer flask at 40% volume of the container. At several times, the flask is microwaved until boiling and swirled to dissolve the agarose. For every 100 mL of agarose solution 5 µL of ethidium bromide is added and swirled. The solution is poured into gel caster and solidifies over time. DNA samples are run in agarose gels with a tracking dye at 130 V for 20 minutes in analytical applications and at 130 V or 150 V for 30-60 minutes when preparing gel extracts. Agarose gel electrophoresis was used to analyze minipreps of bacterial plasmids as from

library colonies. Restriction endonucleases were used by manufacturer protocol to cut the bacterial plasmids to defined sizes and analyze by agarose electrophoresis.

References

- (1) Brustad, E.; Bushey, M. L.; Lee, J. W.; Groff, D. *Angew. Chem. Int. Ed. Engl.* **2008**, *47*, 8220-8223
- (2) Liu, C. C.; Brustad, E.; Mills, J. H.; Groff, D.; Smider, V. V.; Schultz, P. G. *J. Am. Chem. Soc.* **2009**, *131*, 9616-9617
- (3) Bandyopadhyay, A.; McCarthy, K.A.; Kelly, M.A.; Gao J. *Nat Commun.* **2015**, *6*, 6561
- (4) Rauch, B. J.; Porter, J. J.; Mehl, R. A.; Perona, J. J. *Biochem.* **2016**, *55*, 618-628
- (5) Wang, F.; Robbins, S.; Guo, J.; Shen, W.; Schultz, P. G. *PloS ONE* **2010**, *5*, 1-6
- (6) Santoro, S. W.; Wang, L.; Herberich, B.; King, D. S.; Schultz, P. G. *Nature Biotechnol.* **2002**, *20*, 1044-1048

Chapter 3: Testing the Yeast Selection System Based on GAL4

Abstract

Genetically encoding UAAs in eukaryotes are done using a *S. cerevisiae* based selection system. This selection system depends on the conditional expression of the GAL4 transcriptional factor to activate a variety of reporter genes in the yeast genome. The bacterial leucyl aaRS/tRNA pair has been evolved to genetically encode a small number of UAAs in eukaryotes. However, this small set demonstrates outstanding structural diversity. Furthermore, many of the engineered LeuRS variants charge their substrate UAA with remarkable efficiency at unusually low UAA concentrations highlighting its potential to be used for more wide-spread genetic code expansion. Here we have attempted to develop a yeast-based selection system for the evolution of the leucyl pair to genetically encode structurally unique UAAs that remain difficult to achieve using existing pair. Systematic optimization of the selection system was performed to achieve proof-of-concept positive and negative selection of LeuRS variants. We further generated a GFP-reporter expression assay to characterize UAA incorporation in yeast cells.

Introduction

To evolve orthogonal UAA-specific tRNA/aaRS pairs in eukaryotes, a *Saccharomyces cerevisiae* hosted selection system is used (1). The selection system is based on the conditional expression of GAL4, which is a transcription factor that has an N-terminal DNA-binding domain and a C-terminal transcriptional activation domain. GAL4 also has 2 DNA-binding enhancers that increase its activity *in vivo* (2). GAL4 binds DNA with a zinc finger and presumably chelates Zn with two pairs of cysteine residues (3). In this selection system, GAL4 is inactivated with two mutations, T44TAG and R110TAG, and is expressed from a plasmid that is known as pGADGAL4-2TAG (4). These two mutations were previously selected in order to generate the best phenotype to perform the selection. Expression of GAL4 is conditional upon the presence of TAG suppression activity from an orthogonal aaRS/tRNA pair, and when it is expressed GAL4 drives the expression of several reporter genes, URA3, HIS3, and lacZ, which are encoded in the yeast genome. Expression of these reporter proteins can be used to drive positive or negative selections for developing desired aaRS/tRNA pair variants through directed evolution. The selection system was previously optimized and was used to engineer the substrate specificity of the bacterial tyrosyl-aaRS/tRNA pair in *S. cerevisiae* (5). Although, not nearly as successful as the corresponding *E. coli* selection system, a number of UAAs have been genetically encoded in eukaryotes using this platform (6). GAL4 is known to activate the reporter genes of URA3, HIS3, and lacZ.

Expression of lacZ, the gene that produces β -galactosidase, enables active or inactive members (or ones with active/successful tRNA/aaRS pairs versus ones with inactive pairs) to be distinguished based on a classic colorimetric assay. HIS3 and URA3 are genes enable relief from histidine auxotrophy and uracil auxotrophy of the yeast cells, thereby permitting them to grow in the absence of histidine or uracil in the minimal growth media, respectively. The HIS3 gene is known to be significantly leaky even with 2TAG sites in GAL4, and so growth on –His medium can continue to some extent. Therefore, a compound called 3-aminotriazole (3-AT), which is a competitive inhibitor of the HIS3 gene product, is used to further enhance the stringency of the selection. In fact, varying 3-AT concentration can be used to systematically modulate the stringency of the selection pressure. The auxotrophy based on URA3 can also be used as a positive selection. In this study, the HIS3 and URA3 auxotrophies were both successfully used at the same time to make what is called a “double positive selection.” To perform the negative selection, where

members which charge canonical amino acids must be removed, a compound called 5-fluoroorotic acid (5-FOA) is added to the growth medium. 5-FOA is converted into 5-fluorouracil, a toxic metabolite, by the action of the gene product of URA3. Therefore, active amber suppressors will lead to the expression GAL4, which will activate URA3, resulting in the conversion of 5-FOA into the toxic metabolite, which kills the cell.

The selection plasmids used in this system are pESC and pGAD. Both plasmids have selectable markers and origins of replication for both yeast and bacteria so that large quantities of plasmid can be produced in bacteria and then shuttled into yeast. However, the additional genetic components make the plasmids larger and both plasmids are relatively bulky and difficult to transform. The pESC and pGAD plasmids share the 2 μ ori for yeast but have the metabolically selectable auxotrophic markers of TRP1 and LEU2, respectively. The pESC plasmid contains the synthetase and tRNA of interest, while the pGAD plasmid contains the GAL4-2TAG gene.

We wanted to adapt the yeast selection system to facilitate the evolution of the bacterial leucyl-aaRS/tRNA pair. This platform has been previously used to genetically encode a small number of UAAs in eukaryotes. However, this small set of UAAs showcase a high structural diversity. Furthermore, many of the engineered LeuRS variants charge their substrate UAA with remarkable efficiency at unusually low UAA concentrations highlighting its potential to be used for more wide-spread genetic code expansion. However, this system has not been used to genetically encode any new UAAs in the last 7 years. We wanted to resurrect this novel to further genetically encode new UAAs in eukaryotes that remain inaccessible to other available platforms.

The M40, L41, Y499, Y527, and H537 sites of the *E. coli* leucyl synthetase were selected for mutagenesis for the library. These residues are around the active site of the synthetase and their mutations were previously demonstrated to evolve mutant leucyl synthetases with unique substrate specificities (5). The M40 site used 4 different sequences to designate mutations into glycine (GGC), alanine (GCA), valine (GTT), or isoleucine (ATT) and the L41 site used 8 different sequences to designate mutations into serine (TCA), proline (CCC), threonine (ACG), glutamic acid (GAG), glutamate (CAA), asparagine (AAC), leucine (CTT), or valine (GTA). The Y499 site was subjected to change into glycine (GGC), serine (TCA), arginine (CGT), alanine (GCT), isoleucine (ATC), leucine (TTG), and valine (GTA). The Y527 site was

subjected to change into alanine (GCT), glycine (GGT), threonine (ACG), leucine (CTT), isoleucine (ATA), and methionine (ATG). The 4 H537 mutagenic primers specify codons for glycine (GGA), phenylalanine (TTT), threonine (ACT), and glutamate (CAA). The overall diversity of the MLYYH library is 5,376.

The LRS normally contains an editing domain contained inside of connective polypeptide 1, which contains the leucine-specific domain that is inserted into the catalytic site and promotes leucine to be recognized instead of similar amino acids (7). The editing domain is a critical site of operations in UAA incorporation systems because its activity is also generally correlated with how well the LRS recognizes the UAA. Yeast and other eukaryotic cells need A-boxes and B-boxes to express the tRNA and required that the 3'-CCA is enzymatically added only after the transcription (8). The expression of the tRNA at the proper level is one of the most critical factors governing the success of the selection system. The SUP4 tRNA expression system is based on surrounding the desired tRNA with the flanking sequences of an endogenous yeast-tyrosyl suppressor gene and has been previously used to modulate tRNAs expression by adding it downstream of a single PGK1 promoter (9). We adapted this system to systematically modulate the tRNA expression to achieve the optimal levels. In this chapter we present our efforts on optimizing this and other parameters to achieve optimal selection conditions to evolve the bacterial leucyl-tRNA synthetase.

Results

First proof-of-concept selection in yeast

Although the leucyl pair has been previously evolved using the yeast based selection system, the details of this platform was absent. We attempted to acquire and characterize the necessary plasmids from the Schultz group, but they were only able to provide us the corresponding selection plasmids encoding the bacterial tyrosyl pair. We validated that these plasmids provide appropriate reporter expression in response in the presence of amber suppression (supporting information). Next we cloned the leucyl-aaRS/tRNA pair into this plasmid system, and obtained no activity at all.

Since the tRNA expression in eukaryotes is often the most challenging, while protein expressions are straightforward, we attempted to alter the tRNA expression system encoded in the tyrosyl plasmid. A SUP4 promoter system was introduced around the leucyl tRNA and two LeuRS variants were separately subcloned. One encodes a highly polyspecific synthetase (polyLRS) that charges long and hydrophobic amino acids (e.g., acetyllysine) that we have developed. The other charges a fluorescent amino acid dansyl alanine (DanRS). polyLRS/1SUP4 and DanRS/1SUP4 constructs were tested under positive selection conditions in the presence or absence of their respective substrates. The DanRS (a synthetase that charges its substrate with high specificity with no background activity) showed expected patterns of survival (no growth in –UAA plate and survival on +UAA plate) when histidine auxotrophy was used as the positive selection marker (Figure 3.1), when both histidine and uracil was used as positive selection marker (Figure 3.3), and when URA3 was used as a negative selection marker (Figure 3.2). This suggests that the use of SUP4 promoter system provides required level of tRNA expression and provides the first proof-of-concept demonstration of successful development of selection conditions for this pair.

Systematically modulating the tRNA expression to achieve effective proof-of-concept selection

Unfortunately however, when the polyLRS was used from the same plasmid system, no selective growth was observed in the presence or absence of 2-aminocaprillic acid (CAP), one of its substrates. This synthetase variant, like most other engineered aaRS, has a low background activity in the absence of the added UAA, and this low activity is sufficient to provide saturating level of GAL4 expression. Since most engineered variants show background activities like the polyLRS (not as stringent as DanRS; which is not

ideal, but these aaRSs are still useful) it is unlikely that this plasmid system would enable the selection of such variants. We sought to systematically modulate the tRNA expression to achieve a level that will allow differential expression of GAL4 in the presence or the absence of the UAA, such that selection of such aaRS could be achieved. To do this, we systematically truncated the tRNA promoter from the 5'-side to generate a series of variants: -36/SUP4, -27/SUP4, -18/SUP4, and -9/SUP4 tRNA. Two variants were made for each construct: either expressing no aaRS or expressing the polyLRS.

The noRS and polyLRS plasmids harboring -36/SUP4, -27/SUP4, -18/SUP4, and -9/SUP4 tRNA expression cassettes were tested under positive selection condition (histidine auxotrophy, with increasing 3-AT gradients). The -27/SUP4 constructs showed the most promise (Figure 3.6). The noRS -27/SUP4 construct did not survive high concentrations of 3AT and its survival was independent of the presence of the UA. However, this growth did not entirely disappear even by 100 mM 3-AT, especially when the yeast cells were plated densely, suggesting the need to do further optimization. The polyLRS -27/SUP4 grew at 3At concentrations up to 25-50 mM and the growth was strongly dependent on the presence of the UAA in the medium (Figure 3.6). However, the allowing the yeast to grow for longer makes these differences get somewhat obscure, as growth is observed in plates both with or without UAA.

This plasmid system was also tested for proof-of-concept negative selection (Figure 3.7). The noRS -27/SUP4 negative selection and the polyLRS -27/SUP4 were grown in media with increasing concentrations of 5FOA in the presence or absence of the UAA. The noRS -27/SUP4 survived upto a concentration of 0.06% 5FOA and its survival was independent of the UAA. In contrast, the polyLRS -27/SUP4 survived upto the highest concentration of 5FOA in the absence of the UAA, but in the presence of the UAA no growth was observed at 0.06% 5FOA or above, indicating a potentially useful concentration range for negative selection.

Development of a Reporter System to Test UAA incorporation in Yeast

Access to a reporter protein that can be conveniently expressed and purified from yeast is needed to verify UAA incorporation efficiency and fidelity. Fluorescent proteins are great candidates for this as they provide a rapid and convenient assessment of protein expression by measuring their unique fluorescence intensity. The sfGFP and eGFP reporters were not successfully expressed from yeast. However, attempts at

expressing their yeast-optimized counterparts yeGFP was successful and the protein could be purified from these cells using a C-terminal polyhistidine tag by immobilized metal-ion affinity chromatography. The resulting proteins were characterized on the LC/MS successfully (supporting information).

A TAG-containing version of this reporter was also created (yeGFP-TAG) and the ability to express this reporter using the polyLRS suppression system was tested and the resulting yields were compared to the expression level of yeGFP-wt. The yeGFP-wt expression was found to be much higher than yeGFP-TAG. Indeed, the yield of wild-type protein was over 20-fold higher than the corresponding amber mutant. When we tested the effect of the AcK concentration on the efficiency of its incorporation, it was found that increasing AcK concentration considerably improves fluorescence, especially when going up to 2 mM AcK (Figure 3.4).

To try to further improve the suppression efficiency and the expression level of the UAA-modified protein, 3 SUP4 tRNAs were added to the pGAD-GPD-yeGFP-TAG reporter plasmid. The expression of the original reporter system, and the new reporter system that has 3 additional tRNA expression cassettes were tested 4 separate times with polyLRS (Figure 3.5). According to the results, the additional 3 tRNAs is not beneficial to the AcK incorporation efficiency. While the inclusion of additional tRNAs increases the background reporter expression in the absence of the UAA, in the presence of the UAA the reporter expression actually is lower.

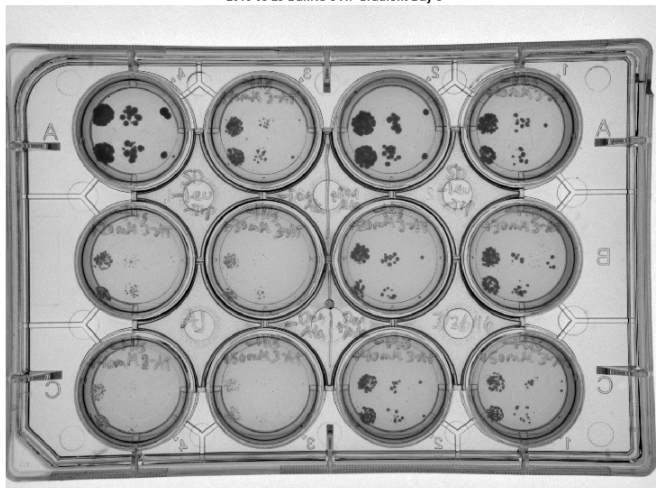


Fig. 3.1 DanRS/1SUP4 3-AT gradient. The *S. cerevisiae* cells with pGADGAL4-2TAG and pESC-DanRS-1SUP4 were spotted in vertical duplicate with 4 μ L $OD_{600} = 0.1, 0.01, \text{ and } 0.001$ (from left to right per well) spots. The left 6 wells of the plate do not have UAA and the right 6 wells of the plate have + 1 mM DanAla. The top left wells of each half contain SD-Leu-Trp medium; the other wells of each half reading from the top left to the bottom right horizontally contain SD-Leu-Trp-His plus a certain amount of 3-AT. The top right wells contain 10 mM 3-AT; the middle left wells contain 20 mM 3-AT; the middle right wells contain 40 mM 3-AT; the bottom left wells contain 40 mM 3-AT; the bottom right wells contain 50 mM 3-AT. The construct grows strongly in all +DanAla wells and more weakly as the concentration of 3-AT increases in the -DanAla wells. Little yeast growth at $OD_{600} = 0.1$ remains at -DanAla 50mM 3-AT condition.



Fig. 3.2 DanRS/1SUP4 5-FOA gradient. The *S. cerevisiae* cells with pGADGAL4-2TAG and pESC-DanRS-1SUP4 were spotted in vertical duplicate with 4 μ L $OD_{600} = 0.1, 0.01, \text{ and } 0.001$ (from left to right per well) spots. The left 6 wells of the plate do not have UAA and the right 6 wells of the plate have +1mM DanAla. The top left wells of each half contain SD-Leu-Trp medium; the other wells of each half, reading from the top left to the bottom right, also contain increasing amounts of 5-FOA. The top right wells contain 0.02% 5-FOA; the middle left wells contain 0.04% 5-FOA; the middle right wells contain 0.06% 5-FOA; the bottom left wells contain 0.08% 5-FOA; the bottom right wells contain 0.10% 5-FOA. The yeast cells grow moderately well in the –DanAla condition at all provided levels of 5-FOA, but only a single colony is left at +0.04% 5-FOA during +DanAla condition and no colonies are observed at higher [5-FOA].



Fig. 3.3 DanRS/1SUP4 double positive gradient. The *S. cerevisiae* cells with pGADGAL4-2TAG and pESC-DanRS-1SUP4 were spotted in vertical duplicate with 4 μ L OD₆₀₀ = 0.1, 0.01, and 0.001 (from left to right per well) spots. The left 6 wells of the plate do not have UAA and the right 6 wells of the plate have +1mM DanAla. The top left wells of each half contain SD-Leu-Trp medium; the other wells of each half reading from the top left to the bottom right also have -His, -Ura, and plus increasing concentrations of 3-AT. The top right wells contain 0.5 mM 3-AT; the middle left wells contain 1 mM 3-AT; the middle right wells contain 2.5 mM 3-AT; the bottom left wells contain 5 mM 3-AT; the bottom right wells contain 10 mM 3-AT. The whole difference of yeast growth is observed in all wells across UAA conditions except for the control wells as expected.

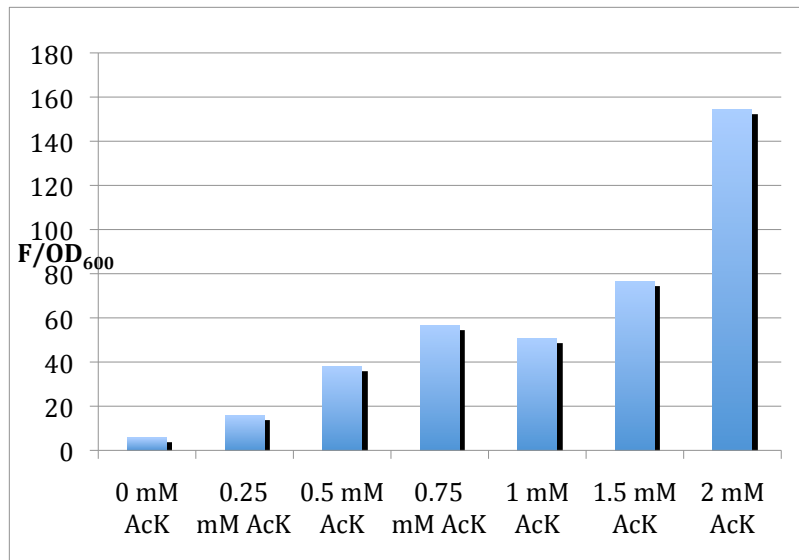


Fig. 3.4. Expression of yeGFP-TAG in *S. cerevisiae* using the polyLRS suppression system in the presence of increasing concentrations of AcK. Yeast cells containing the pGAD-GPD-yeGFP-TAG plasmid were induced, and allowed to express the reporter in the presence of increasing concentrations of AcK, and their lysates were analyzed for yeGFP expression by fluorescence. In general, fluorescence increases as the amount of AcK used increases in this range of concentrations.

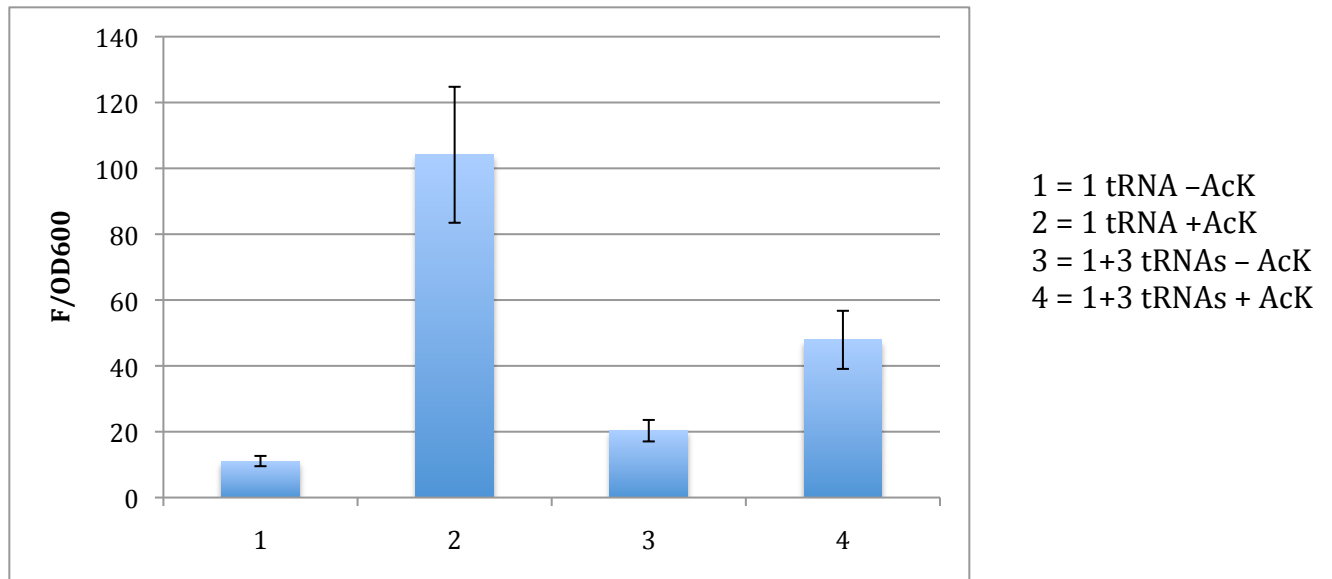


Fig. 3.5 Average of four trials of *S. cerevisiae* yeGFP expression with normal reporter plasmid against its counterpart harboring 3 additional tRNAs. The fluorescence of +/-AcK lysates from the normal reporter system and a reporter system with 3 additional SUP4 tRNAs on the pGAD plasmid were tested for reporter expression by fluorescence. The additional tRNAs cause higher background and lesser activity according to the experiments.

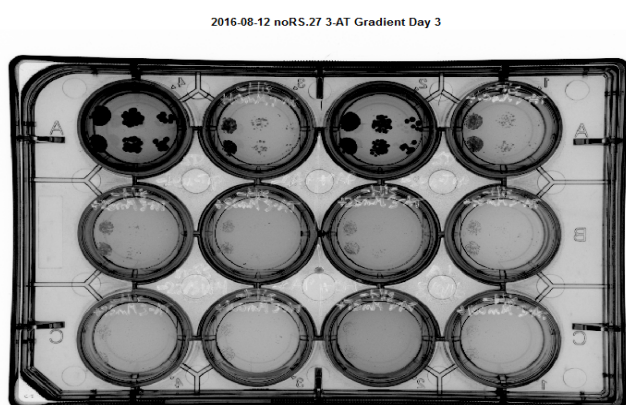
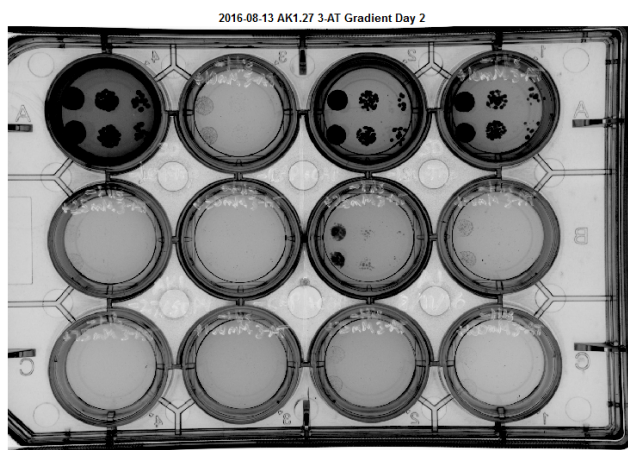


Fig. 3.6. Proof-of-concept positive selection using AK1 and noRS constructs harboring the truncated SUP4 tRNA expression cassette. Top: *S. cerevisiae* with pGADGAL4-2TAG and pESC-AK1-(-27)/SUP4 was plated by 4 μ L of OD₆₀₀ = 0.1, 0.01, and 0.001 (from left to right) spots in vertical duplicate per well. The left 6 wells contain no UAA and the right 6 wells contain 1mM CAP. The top left wells contain SD-Leu-Trp; the other wells contain SD-Leu-Trp-His plus increasing concentrations of 3-AT. Per half of plate, the top right wells, middle left wells, middle right wells, bottom left wells, and bottom right wells contain 10 mM, 25 mM, 50 mM, 75 mM, and 100 mM 3-AT respectively. Bottom: Analogous 12-well 3-AT gradient for the pESC-noRS-(-27)/SUP4 construct. The AK1 synthetase started showing correct differences based on CAP presence or absence at intermediate concentrations of 3-AT. The noRS background with the same tRNA is small and generally limited to the spots with the higher OD values.



Fig. 3.7. Proof-of-concept negative selection using AK1 and noRS constructs harboring the truncated SUP4 tRNA expression cassette. *S. cerevisiae* with pGADGAL4-2TAG and pESC-AK1-(-27)/SUP4 was plated by 4 μ L of OD₆₀₀ = 0.1, 0.01, and 0.001 (from left to right) spots in vertical duplicate per well. The left 6 wells contain no UAA and the right 6 wells contain 1mM CAP. The top left wells contain SD-Leu-Trp; the other wells also contain increasing concentrations of 5-FOA. The top right wells, middle left wells, middle right wells, bottom left wells, and bottom right wells contain 0.02%, 0.04%, 0.06%, 0.08%, and 0.10% 5-FOA, respectively. Bottom: Analogous 12-well 5-FOA gradient for noRS version of construct. The AK1 version grew well on all wells of -CAP and stopped growing at 0.06% 5-FOA of +CAP. The noRS version was being impeded at 0.08% 5-FOA and greater.

Discussion

In this chapter we have systematically optimized the yeast based selection system to evolve the leucyl pair for genetically encoding new UAAs in eukaryotes. As we have shown the rationally developed polyLRS is capable of charging AcK, an important post-translational modification into proteins expressed in mammalian cells. The ability to further fine-tune this synthetase would further improve the incorporation efficiency of this UAA. Given the plasticity of the active site of this aaRS, many other UAA-specific aaRS may also be evolved using this system.

Although a platform to evolve this pair was developed previously, no sequence information regarding those selection plasmids are unavailable. We acquired the corresponding tyrosyl selection system from the Schultz lab and replaced the tyrosyl pair with the leucyl pair, but the resulting system was inactive. We then substituted the tRNA expression cassette with a stronger SUP4 promoter and showed that resulting plasmid was too active, and was unable to exhibit the differential growth pattern in the presence or absence of the substrate UAAs that is required for selection. To find the right balance of tRNA expression, we then systematically truncated the SUP4 promoter on the tRNA. One such truncated constructs was able to provide the optimal level of tRNA expression leading to the differential growth pattern in the presence or absence of the substrate UAAs that is required for selection. These results establish a potentially powerful selection system to genetically new UAAs in eukaryotes based on the leucyl pair. We also have generated a library of mutants where key residues of the LeuRS active sites were randomized. Finally we have also established a convenient fluorescent protein based reporter system to validate efficiency and fidelity of UAA incorporation using the yeast expression system.

Materials & Methods

General

Cells were imaged on ChemiDoc MP system. Fluorescence values were taken on SpectraMax M5 plate reader. Absorbance values were taken on Nanodrop 2000c Spectrometer. Primers and gBlocks were purchased from IDT; restriction enzymes were purchased from NEB. SD-Leu-Trp-His-Ura medium was purchased in pouches from Clontech and adjusted as needed with amino acids and selection reagents depending on trial. Amino acids were purchased from Sigma-Aldrich. 3-AT was purchased from TCI America. 5-FOA and Zymoprep were purchased from Zymo Research. Y-PER was purchased from Thermo Scientific.

MLYYH LRS Library Generation

The library was cloned in two different rounds of mutagenesis. The first round of mutagenesis affected the M40 and L41 sites and the second round of mutagenesis was done for the Y499, Y527, and H537 sites on the corresponding 2-residue mutant synthetase. The template of the first round of mutagenesis was a polyLRS with the T252A mutation of the editing domain. Later, an additional library was constructed with the original T252 residue because of findings regarding the efficiency of this mutation in the selection system. For the first round of mutagenesis, the synthetase was amplified in 2 overlapping fragments via PCR. The first fragment was amplified by a forward terminal primer and a mixture of 32 mutagenic reverse primers that flank both sides of the M40-L41 site. The 32 different combinations of those amino acid options were ordered on 32 different reverse primers from Integrated DNA Technology. The second fragment was amplified by a forward primer that sits on the downstream flank of the M40-L41 site and a reverse terminal primer. After amplifying and purifying the two fragments, an overlap extension PCR was done between the two fragments without any primers. The two fragments were generally added to the reaction mixture at a molar ratio based on 1 µg of the larger fragment per 50 µL reaction. The OE PCRs were analyzed by agarose gel electrophoresis before gel purification of the product-sized band. This DNA was used as the OE template for the amplification of whole mutant synthetases in a series of downstream PCRs. After large amounts of whole synthetases are produced, they are digested on microgram scale by the restriction endonucleases EcoRI-HF and NotI-HF based on 1 unit of each enzyme per 1 µg of DNA in the

reaction. Large amounts of insert are reacted with large amounts of vector in 100 μ L ligation reactions with 8 μ L of T4 DNA ligase.

For the second round of mutagenesis, the leucyl synthetase template was the 2-site leucyl library generated from the first round. This template was amplified into 3 different fragments, each of which carried 1 of the 3 mutations. The Y499 site was amplified by a terminal forward primer and 7 mutagenic reverse primers specifying 1 codon each for the 7 selected residues. The Y527 site was amplified by a forward primer that overlaps with the downstream flank of Y499's reverse primer and 6 mutagenic reverse primers that specify codons for each targeted residue. The H537 site was amplified by a forward mutagenic primer that is located on the downstream flank of Y527's reverse primer and a terminal reverse primer. The first OE PCR was done between the Y527 and H537 fragments. After gel purification and rounds of amplification from the OE template, the Y499 fragment was contributed to the Y527-H537 fragment in a second OE PCR. After amplifying the whole synthetase with additional rounds of PCR, the complete library was cloned into vectors with the 1SUP4, and -27/SUP4 tRNA constructs at different times. The -27/SUP4 tRNA vectors were cloned with A252 and T252 versions of the library.

X-Gal Agarose Overlay

A solution of 200 mM sodium phosphate (Na_2HPO_4) is dissolved in a volume of ddH₂O that is about 90% of the final projected volume of solution. This solution is adjusted with HCl at a pH probe to a pH of 7.0-7.4. Then 1% agar (w/v) is added to the solution and the whole thing is autoclaved. During autoclaving, chloroform is retrieved from the bottle with a syringe and transferred into a glass bottle. By the Bunsen burner flame, with a plastic pipette, the chloroform is dripped onto the colonies of each subject plate. The chloroform is allowed to evaporate over time. Chloroform is added and evaporated in this way several times. After autoclaving, X-Gal is added to the buffered agar solution. 10 mL of this X-Gal solution is added to every chloroform-lysed plate and allowed to solidify next to the flame. Overlay plates are incubated upside-down at 37 °C to develop color.

Yeast Miniprep Protocol

Grow 1.5 mL of subject yeast in selective SD medium overnight. Large colonies will inoculate culture tubes with visible OD_{600S} and will be saturated overnight, but small and ordinary colonies may take longer to reach saturation. Pipet 1.5 mL of culture of 1.5 mL Eppendorf tube. Pellet by centrifuging at

11,000 x g for 20 seconds and remove the supernatant. Resuspend each pellet in 250 μ L zymolyase solution. Incubate the resuspensions at 37 °C for 30 minutes. Add 200 μ L of lysis solution and invert to mix. Add 100 μ L of KAc solution and immediately invert to mix. Centrifuge at 15,000 x g for 10 minutes and transfer the supernatants to new tubes. Add 700 μ L isopropanol to each tube. Mix by inversion and centrifuge at 15,000 x g for 5 minutes. Remove supernatant and add 300 μ L PB1. Incubate at 37 °C for 30-60 minutes; do not resuspend manually or with pipette. Add 300 μ L isopropanol and centrifuge at 15,000 x g for 5 minutes. Remove supernatant and rinse pellet with 1 mL ethanol a few times. Air dry with caps open for 15 minutes. Resuspend final pellet in about 50 μ L TE or other storage buffer. Growth and zymolyase steps were unchanged throughout the investigation. The KAc inversions were done individually from ice for each tube, immediately after addition, in order to control for faulty precipitations. The centrifugation and transfer of the precipitation can be done 1-2 times.

Yeast Miniprep Reagents

The “KAc” solution is a 5 M potassium acetate solution neutralized to pH 5.5 with 5 M acetic acid. The lysis solution contains 50 mM tris-HCl pH 8.0, 50 mM EDTA, and 1.2% SDS. The zymolyase solution is prepared in 92% volume with 1.2 M sorbitol, 10 mM tris, and 10 mM CaCl_2 . This solution is adjusted in terms of pH to 8.0 and gets autoclaved. Immediately before addition of zymolyase solution, for every 920 μ L of 92% volume solution, add 70 μ L 10 mg/mL zymolyase enzyme (stored in aliquots at -80 °C) and 10 μ L β -mercaptoethanol. Keep tube open as little as possible after adding BME. Mix carefully by gentle inversions and quickly spin the contents to the bottom of the tube. In this protocol, the exact amount of required zymolyase solution is created and used. BME, SDS, tris, acetic acid, and sorbitol were purchased from Fisher Scientific. CaCl_2 , EDTA, and potassium acetate were purchased from Sigma-Aldrich. Zymolyase was purchased from US Biologicals.

GAL4 Cloning Tests

To revert the 2 TAG sites of GAL4 back to wild-type sequence for testing, I used two mutagenic reverse primers (one for each site). The first fragment (TAG1) was amplified with the terminal forward primer and the first mutagenic reverse primer. The second fragment (TAG2) was amplified with a forward primer that sits on downstream flank of TAG1’s reverse primer and a mutagenic reverse primer. The third fragment (C-terminal) was amplified with a forward primer that sits on the downstream flank of TAG2’s

reverse primer and a terminal reverse primer. Each fragment was amplified and gel purified. GAL4 used a combination OE PCR where the C-terminal and TAG2 fragments were thermocycled for 12 cycles and then the TAG1 fragment was added to the reaction mixture to thermocycle for 12 more cycles before final extension. This OE PCR directly produced the full-length GAL4-wt. To make GAL4-1TAG, there were 2 different fragments: one fragment is the same as the C-terminal fragment of the other cloning, and the other is amplified from the terminal forward primer and TAG2's mutagenic reverse primer. This procedure means that the 2nd TAG site in GAL4 is reverted to the wild-type sequence. After OE PCR and amplification, the GAL4-1TAG is ready to be cloned.

SUP4 Truncation Cloning

The SUP4 gBlock was amplified with 4 forward primers that contain an AgeI site, NheI site, and sequence recognitions of only the first 36, 27, 18, and 9 base pairs of 5'-SUP4 relative to the tRNA. Each construct was amplified by PCR and digested with AgeI and AvrII to make inserts that are entered into AgeI/NheI pESC vectors.

Yeast Transformation Protocols

Inoculate single colony of subject yeast into 5 mL selective medium. Grow at 30 °C overnight. Passage into a fresh 5 mL medium by transferring 100 µL and growing at 30 °C overnight. Repeat passaging. Measure OD₆₀₀ of final starter culture. Passage to 50 mL culture with OD₆₀₀ = 0.2. Grow at 30 °C with shaking to reach OD₆₀₀ = 1.5. Transfer to 50-mL conical tube. Pellet at 2,000 x g for 5 minutes at 4 °C. Decant supernatant and add 25 mL 100 mM sterile lithium acetate. Resuspend by shaking and add 250 µL fresh 1 M DTT. Loosen and tape cap; incubate at 30 °C with shaking for 10 minutes. Move cells to ice and centrifuge at 2,000 x g for 5 minutes at 4 °C. Decant and add 25 mL ice-cold sterile ddH₂O. Wash cells and then centrifuge at 2,000 x g for 5 minutes at 4 °C. Add ~100 µL ddH₂O to final pellet and immediately proceed to electroporation. Electroporate 100 µL of yeast cells in 1-mm cuvette at 0.25 kV. Add 1 mL SD medium 2 times and contribute to culture tube. Incubate culture tube at 30 °C for 1 hour to recover. (10)

For chemical transformation, grow 5 mL of appropriate strain in selective SD medium overnight at 30 °C with shaking. Use this culture to inoculate 100 mL selective SD medium to OD₆₀₀ = 0.1 and grow at 30 °C with shaking until OD₆₀₀ = 0.4 is reached. Transfer these samples into sterile 50-mL conical tubes and centrifuge at 1,500 x g for 5 minutes at 4 °C. Discard supernatant and add 25 mL sterile ddH₂O to each

tube. Resuspend the cells into the water and centrifuge again at 1,500 x g for 5 minutes at 4 °C. Discard supernatant and resuspend in 5 mL 15% glycerol per tube. Combine samples across tubes and centrifuge under same conditions again. Discard supernatant and resuspend in 900 µL 15% glycerol per 100 mL culture. Store competent cells in 100 µL aliquots at -80 °C.

To transform chemically (11), thaw and centrifuge the appropriate number of yeast competent cells at 9,000 x g for 1 minute. Heat an aliquot of salmon sperm DNA at 100 °C in PCR machine for 5 minutes. Keep cells and DNA on ice otherwise. For every transformation mixture, add 36 µL 1 M lithium acetate, 10 µL 10 mg/mL salmon sperm DNA, 1 µg plasmid DNA, 240 µL 50% PEG-3350, and volume of ddH₂O to reach a final 1X volume of 360 µL. Mix solution well with pipette and add entire solution to 1 competent cell pellet. Vortex the pellet into the solution for 10-20 seconds. Heat shock the transformation tubes in a PCR machine at 42 °C for 40-60 minutes. Every 5 minutes of heat shock, take out tubes and invert them to mix. After the heat shock, centrifuge the tubes at 3,000 x g for 5 minutes. Pour and pipette to remove supernatant and resuspend in 500 µL sterile ddH₂O per tube. Generally, 100 µL of this solution is plated onto an appropriate SD agar plate.

Yeast Spot Plating

The appropriate strains of *S. cerevisiae* were grown overnight in 2 mL SD-Leu-Trp at 30 °C with shaking after inoculation of 1 large colony from respective SD-Leu-Trp plate. In the morning, 1 of 2 cultures for each strain was inoculated with 1 mM of the respective UAA and all cultures were grown for an additional 4 hours at 30 °C with shaking. The cells were pelleted by centrifugation at 9,000 x g for 1 minute. The medium was then removed with a sterile pipette tip and the cells were washed with 1 mL of PBS or 0.9% NaCl each. The cells were centrifuged as before and the washing step was repeated. After two washes, the cells were pelleted under the same conditions again and then 1 mL of PBS or 0.9% NaCl was added. The cells stood at room temperature for 30 minutes to equilibrate in the solution. The cells were pelleted under the same conditions again and then washed two more times with 1 mL PBS or 0.9% NaCl each. After the last wash, the cells were resuspended in 1 mL PBS or 0.9% NaCl and these were the solutions that were measured for OD₆₀₀. The cells were then divided into dilution series. The amount of cells to make 1 mL of OD₆₀₀ = 0.1 solution was calculated and transferred into a tube, with the diluent to 1 mL being PBS or 0.9% NaCl. Afterwards, the other members of the dilution series were made by

transferring 100 μ L of the previous solution into 900 μ L of fresh PBS or 0.9% NaCl the respective number of times. For mock selections, the cells were co-inoculated from the dilution series samples with the concentrations matching the desired cellular ratios. The cells were thoroughly mixed with the pipette tip in each solution during each transfer. To make the spot plate, by the Bunsen burner, a 4-5 μ L aliquot of each 1 mL tube was carefully dripped with a pipette into the suggested position of each well the respective number of times (2x per well for duplicate) in series. For 6-well plates, a 10 μ L aliquot was used instead. After all drops were spotted onto the plate, the lids were turned over and the spots were left to dry by the Bunsen burner for about 30 minutes or until done. The plates were then incubated upside-down at 30 °C to grow for 2-4 days.

Design of Yeast Reporter Plasmids

To accompany a selection system, there should also be plasmids that are designed to produce noticeable phenotypes in response to the intensity and efficiency of the UAA incorporation. After unsuccessfully trying to correlate the fluorescence signal of eGFP-wt (12), sfGFP-151-TAG and sfGFP-3-TAG to the activity of polyLRS or wtLRS synthetases, the cloning moved to the codon-optimized yeGFP (yeast enhanced green fluorescent protein). Because yeGFP (13) contains the degenerate codons for each residue that are more common and more facile in yeast cells, yeGFP is translated much more efficiently than eGFP. The pGADGAL4-2TAG plasmid was used as the starting point for the cloning exercises. The GPD promoter was inserted in place of the ADH1 promoter and GAL4 gene in the plasmid with the SphI-HF and EcoRI-HF restriction sites. Afterwards, the yeGFP was separately cloned downstream of the GPD promoter with SpeI-HF and KpnI-HF restriction sites. The terminator remained ADH1. The GPD promoter was selected because its strength was known to be very strong. After the pGAD-GPD-yeGFP plasmid was constructed, it was used as the template to perform a G4TAG mutagenesis of yeGFP. The yeGFP was mutated with a reverse primer on a fragment that included some of the upstream base pairs from GPD promoter and an overlapping second fragment was amplified from the remainder of yeGFP. The two fragments were fused by OE PCR and the resulting template was amplified by PCR before being digested with the SpeI-HF and KpnI-HF to be added to the respective sites of a pGADGPD vector.

The 100 mL yeast cultures of pESC-polyLRS-1SUP4 and pGAD-GPD-yeGFP or pGAD-GPD-yeGFP-TAG were grown at 30 °C with shaking. Each culture was pelleted and then 10 mL of a cocktail of

Y-PER, 1x Halt protease inhibitor, Pierce universal nuclease, and 100 mM DTT was added. After nutation for 30 minutes, the yeast cells were pelleted and their supernatants were purified by Ni-NTA chromatography according to manufacturer protocol. The purified proteins were analyzed by LC/MS. For expression studies, 1 mL of the desired yeast strain was inoculated into SD-Leu-Trp+/-UAA to grow overnight at 30 °C with shaking. A specified number of yeast cells (such as $OD_{600} = 0.3$ or $0.4 @ 1 \text{ mL}$) were aliquoted into separate tubes and pelleted. The cells were lysed in the Y-PER cocktail as before and then a dilution of the lysate (10-fold) was added to PBS in a 96-well plate. The fluorescence of the 96-well plate was recorded with excitation of 488 nm and emission of 514 nm.

References

- (1) Chin, J. W.; Cropp, T. A.; Anderson, C.; Mukherji, M.; Zhang, Z.; Schultz, P. G. *Sci.* **2003**, *301*, 964-967
- (2) Mitzutani, A.; Tanaka, M. *EMBO J.* **2003**, *22*, 2178-2187
- (3) Johnston, M.; Dover, J. *Gen.* **1988**, *120*, 63-74
- (4) Chin, J. W.; Cropp, T. A.; Chu, S.; Meggers, E.; Schultz, P. G. *Chem. & Biol.* **2003**, *10*, 511—519
- (5) Cropp, T. A.; Anderson, J. C.; Chin, J. W. *Nature Protocols* **2007**, *2*, 2590-2600
- (6) Wiltshi, B. *Fung. Gen. and Biol.* **2016**, *89*, 137-156
- (7) Yan, W.; Tan, M.; Eriani, G.; Wang, E. *Nucleic Acids Res.* **2013**, *41*, 4988-4998
- (8) Wang, Q.; Wang, L. *Methods Mol. Biol.* **2012**, *794*, 199-213
- (9) Chen, S.; Schultz, P. G.; Brock, A. *J. Mol. Biol.* **2007**, *371*, 112-122
- (10) Van Deventer, J. A.; Kelly, R. L.; Rajan, S.; Wittrup, K. D.; Sidhu, S. S. *Protein Eng. Des. Sel.* **2015**, *28*, 317-325
- (11) Gietz, R. D.; Schiestl, R. H. *Nat. Protocols* **2007**, *2*, 38-41
- (12) Arpino, J. A. J.; Rizkallah, P. J.; Jones, D. D. *PLoS ONE* **2012**, *7*, 1-8
- (13) Kaishima, M.; Ishii, J.; Matsuno, T.; Fukuda, N.; Kondo, A. *Sci. Rep.* **2016**, *6*, 1-15

Chapter 4: Future Directions

Use the Right Crystal Structure of pBoF Synthetase

A more precise crystal structure of the pBoF synthetase rather than the corresponding wild-type *M. jannaschii* tyrosyl synthetase might clarify a more accurate path to the rational design of a successful evolution experiment to genetically encode AcpBoF. If the pBoF synthetase actually had significant structural differences at the active site, then the rational design done in this study could have been skewed. Another approach would be to start from the scratch and build a large naïve library to directly screen for AcpBoF. It is possible that pBoF and AcpBoF are treated differently by the host translation machinery or other cellular factors. The amine reactive AcpBoF may stick to a variety of proteins in the cell, which may substantially deplete its effective intracellular concentration, or perturb its cellular uptake. These aspects must be carefully considered.

Yeast Selections with Leucyl Synthetase

The platform of performing the leucyl synthetase selection inside of *S. cerevisiae* appears to be ready, providing convincing proof-of-concept demonstration of positive and negative selection with both polyLRS and DanRS. One concern with the selection system is the lack of desired level consistency, which may be improved through further optimization of this system. The library of LeuRS variants which should be very rich in potential hits can also be used to truly validate the selection system by evolving novel UAA-specific variants. The selection with the leucyl synthetase is expected to target N-acetyl-lysine first, but can also feature various other unnatural amino acids once the system is validated.

Modulation of Selection Activity with Origins of Replication

Testing should be done with respect to the copy numbers of pESC and pGAD plasmids inside of yeast cells. The origins of replication (oris) control the number of plasmids in a provided yeast cell, and the pESC and pGAD plasmids used in this study only employed the 2 μ ori, which is a high copy number episomal one. The interaction of two different plasmids with the same 2 μ ori can have been a liability that might have affected the results of selection. Employing different yeast oris, such as low copy number centromeric ones, could be used to evaluate its effect on the selection system.

Change the TAG Sites of the GAL4 Gene

The T44TAG and R110TAG sites were previously selected because they appeared to show the best selective phenotypes during the early studies. Because the different aaRS/tRNA pairs appear to have different suppression preferences, it might be possible to improve the stringency of the selection system by changing the TAG sites of the GAL4 gene. It is also possible that having additional TAG sites can improve the operations of the selection since one of the persistent problems I faced comes from the leaky nature of the positive selection. A more stringent GAL4 activation by the synthetase/tRNA pair might be able to fix the problems with the selection.

Appendix 1: Supporting Information

Pictures of Bacterial Selection Plates

Mock Selection –Cm -pBoF

2015-03-22 Mock Selection -CHLOR -pBoF (Washed, Less CHLOR, More Colonies)



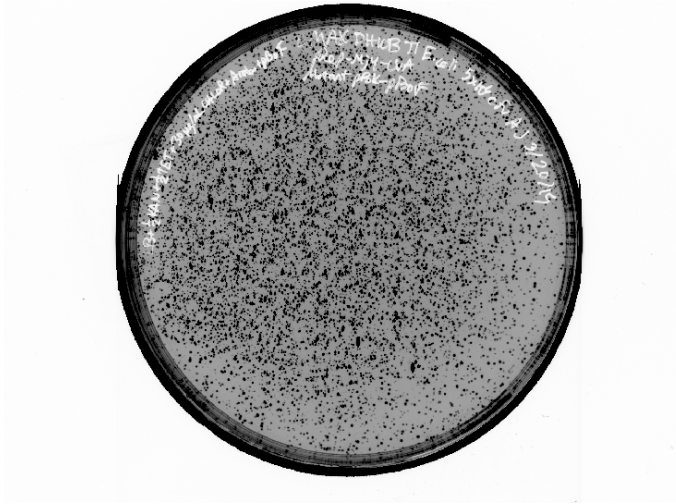
Mock Selection +Cm -pBoF

2015-03-22 Mock Selection +CHLOR -pBoF (Washed, Less CHLOR, More Colonies)



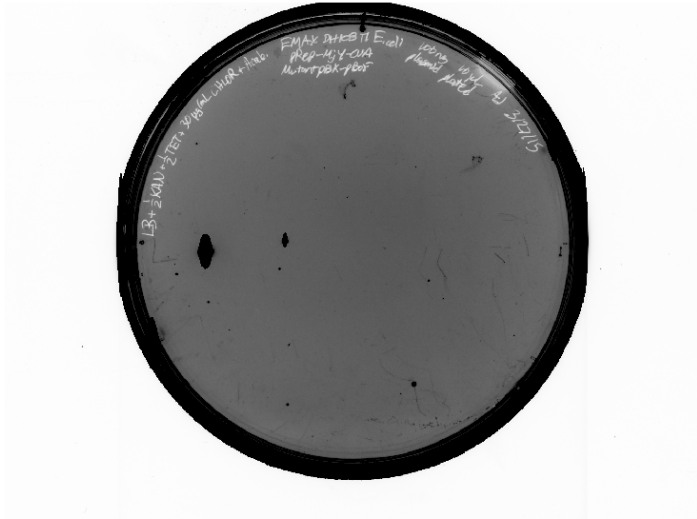
Mock Selection +Cm +pBoF

2015.03.22 Mock Selection +CHLOR +pBoF (Washed, Less CHLOR, More Colonies)

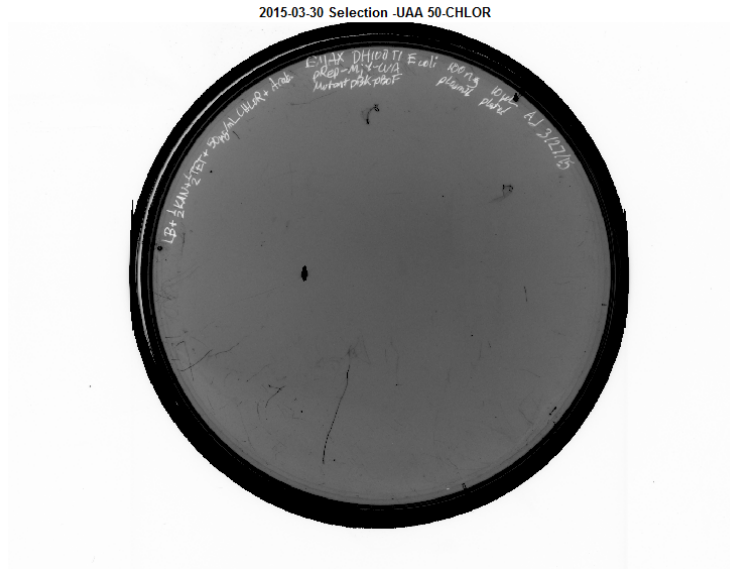


First Selection +30 µg/mL Cm -UAA

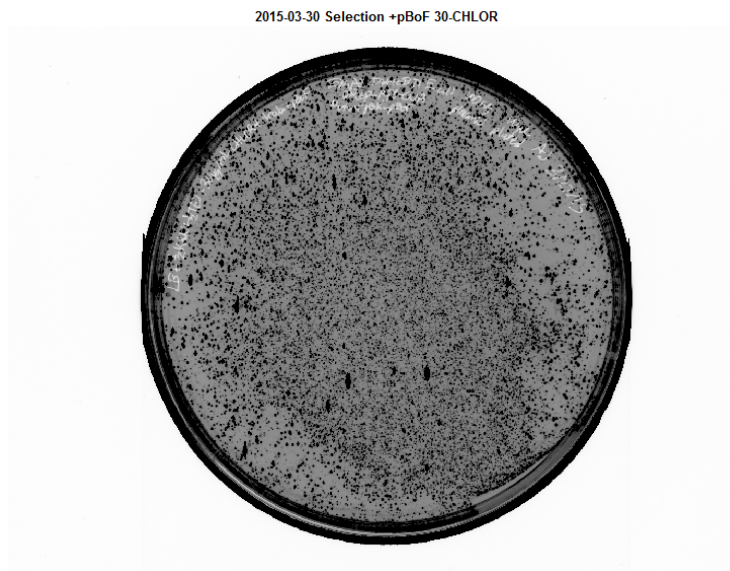
2015.03.30 Selection -UAA 30.CHLOR



First Selection +50 µg/mL Cm -UAA



First Selection +30 $\mu\text{g/mL}$ Cm +pBoF



First Selection +50 $\mu\text{g/mL}$ Cm +pBoF

2015-03-30 Selection +pBoF 50-CHLOR

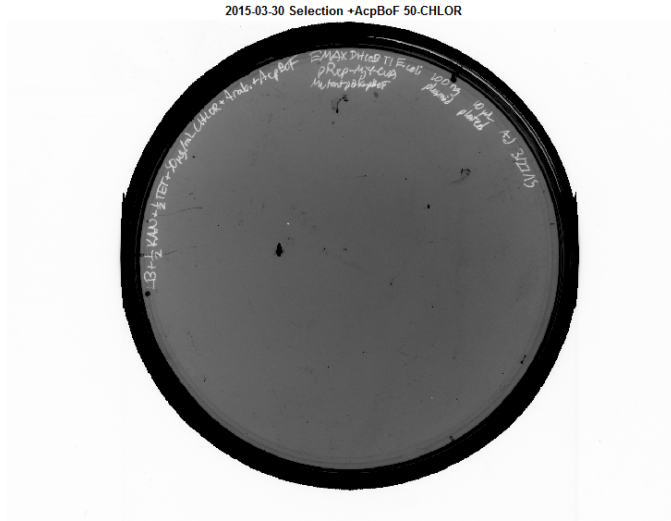


First Selection +30 $\mu\text{g/mL}$ Cm +AcpBoF

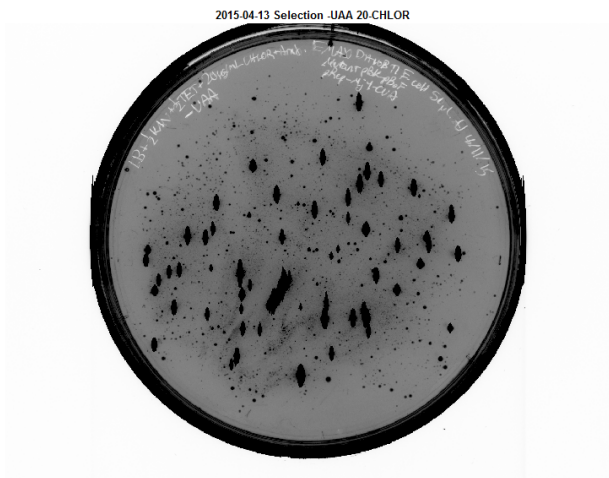
2015-03-30 Selection +AcpBoF 30-CHLOR



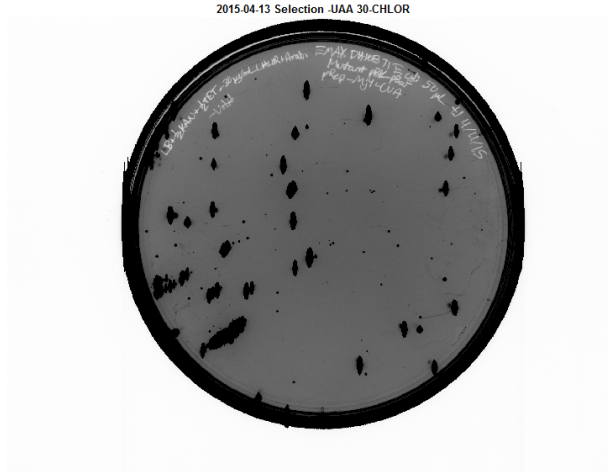
First Selection +50 µg/mL Cm +AcpBoF



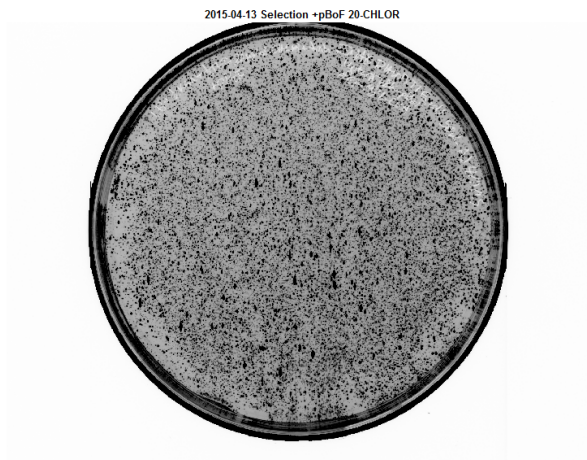
Second Selection +20 µg/mL Cm -UAA



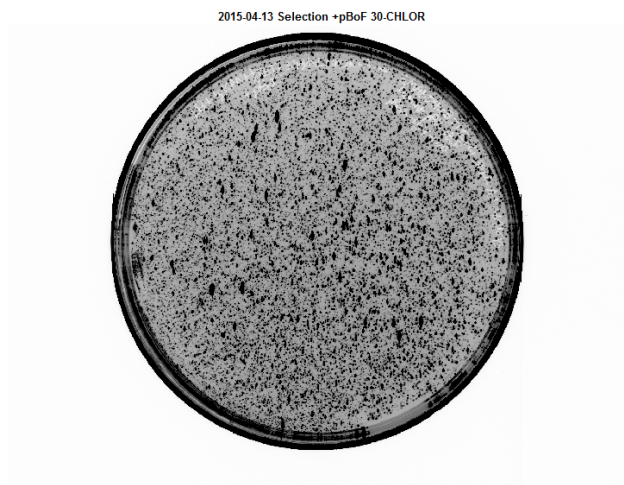
Second Selection +30 $\mu\text{g/mL}$ Cm -UAA



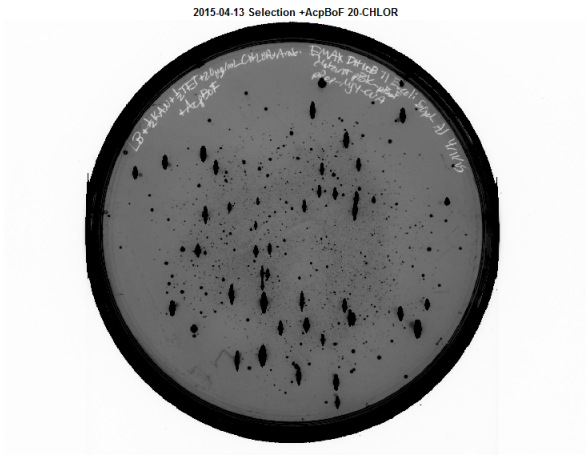
Second Selection +20 $\mu\text{g/mL}$ Cm +pBoF



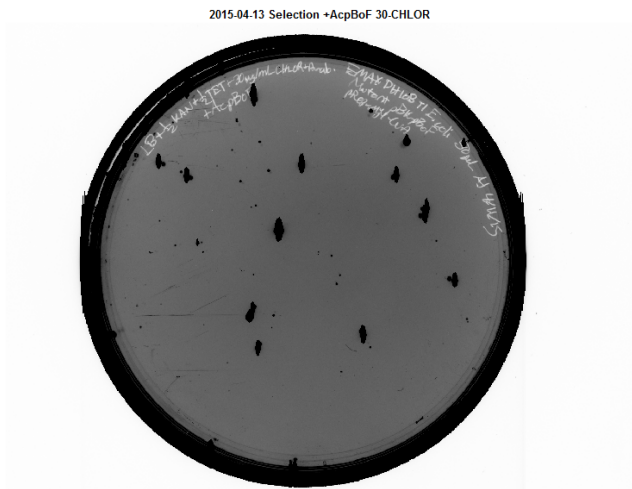
Second Selection +30 $\mu\text{g/mL}$ Cm +pBoF



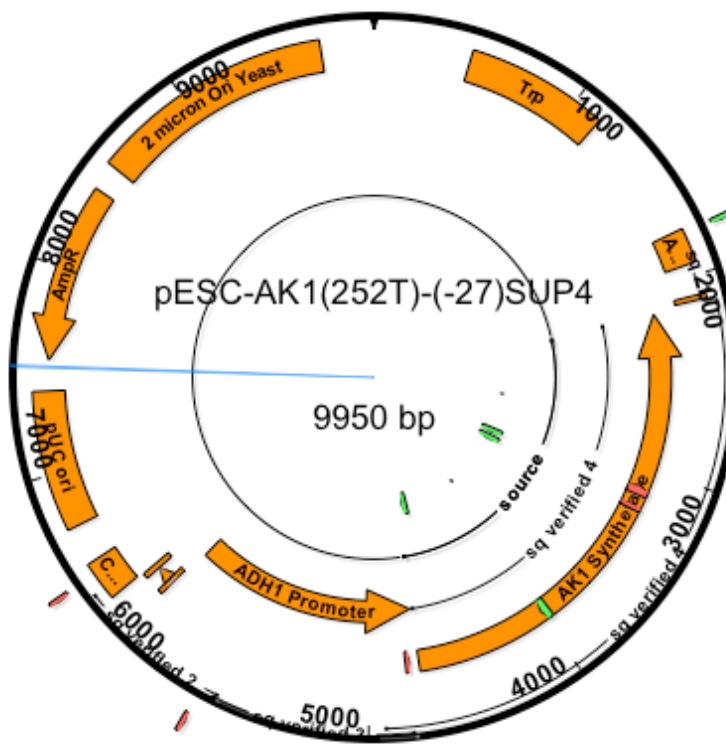
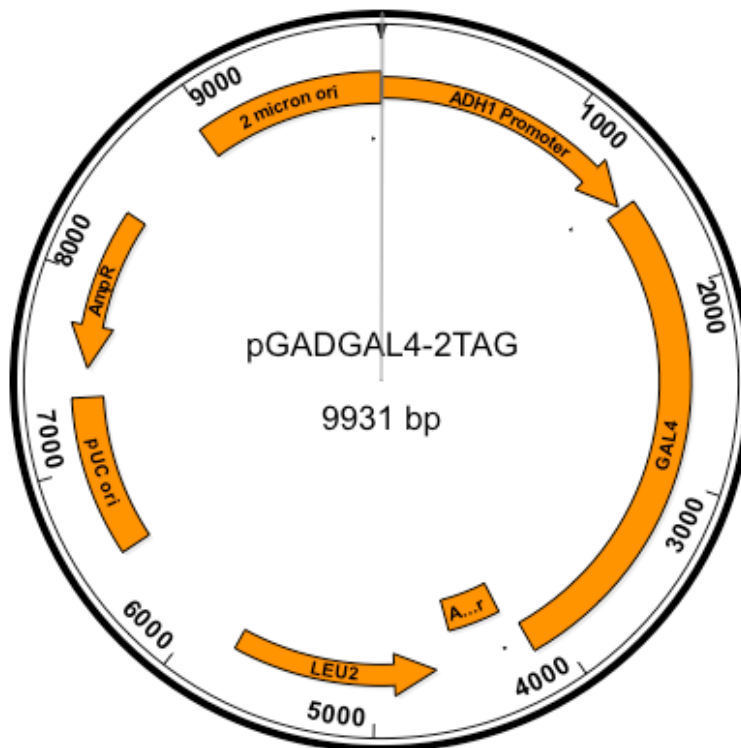
Second Selection +20 µg/mL Cm +AcpBoF

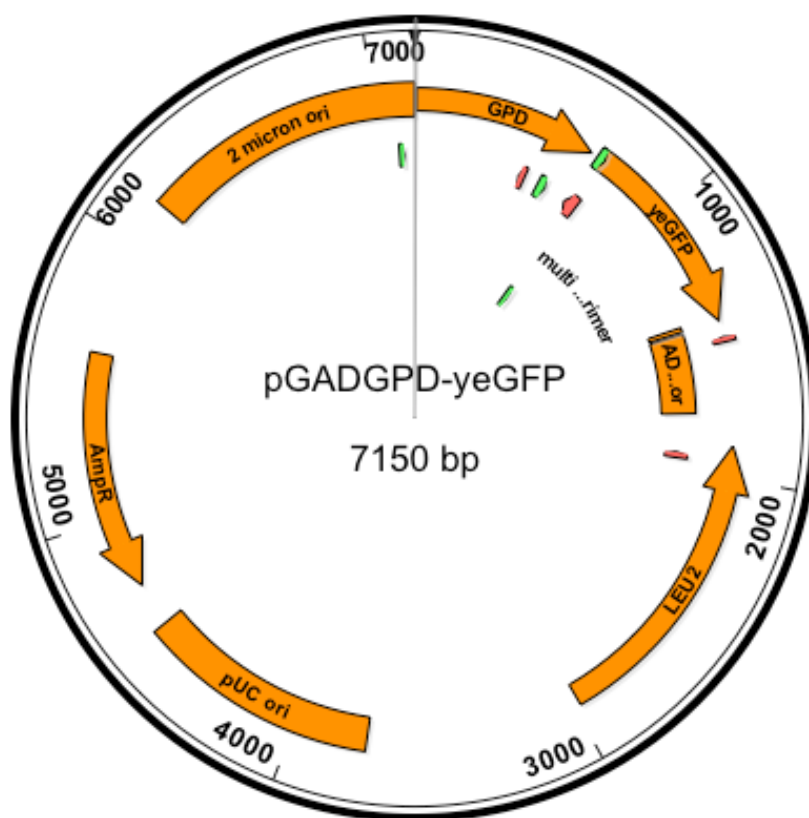


Second Selection +30 µg/mL Cm +AcpBoF



Selected Yeast Nucleotides





Primer Set for ML-Site Mutagenesis

Mutagenic sites are underlined.

Residue Combination	Primer Sequence (5' => 3')
Glycine/Serine	GTGTAGTCGACCAGAAGGATAGGGGGAGCCAGACAGGCAGTAATACTTCTCTTTGCTCTC
Alanine/Serine	GTGTAGTCGACCAGAAGGATAGGGGGATGCAGACAGGCAGTAATACTTCTCTTTGCTCTC
Valine/Serine	GTGTAGTCGACCAGAAGGATAGGGGGAAACAGACAGGCAGTAATACTTCTCTTTGCTCTC
Isoleucine/Serine	GTGTAGTCGACCAGAAGGATAGGGGGAAATAGACAGGCAGTAATACTTCTCTTTGCTCTC
Glycine/Proline	GTGTAGTCGACCAGAAGGATAGGGGGGCCAGACAGGCAGTAATACTTCTCTTTGCTCTC
Alanine/Proline	GTGTAGTCGACCAGAAGGATAGGGGGGTGCAGACAGGCAGTAATACTTCTCTTTGCTCTC
Valine/Proline	GTGTAGTCGACCAGAAGGATAGGGGGGAACAGACAGGCAGTAATACTTCTCTTTGCTCTC
Isoleucine/Proline	GTGTAGTCGACCAGAAGGATAGGGGGGAATAGACAGGCAGTAATACTTCTCTTTGCTCTC
Glycine/Threonine	GTGTAGTCGACCAGAAGGATAGGGCGTGCCAGACAGGCAGTAATACTTCTCTTTGCTCTC
Alanine/Threonine	GTGTAGTCGACCAGAAGGATAGGGCGTTGCAGACAGGCAGTAATACTTCTCTTTGCTCTC
Valine/Threonine	GTGTAGTCGACCAGAAGGATAGGGCGTAACAGACAGGCAGTAATACTTCTCTTTGCTCTC
Isoleucine/Threonine	GTGTAGTCGACCAGAAGGATAGGGCGTAATAGACAGGCAGTAATACTTCTCTTTGCTCTC
Glycine/Glutamic Acid	GTGTAGTCGACCAGAAGGATAGGGCTCGCCAGACAGGCAGTAATACTTCTCTTTGCTCTC
Alanine/Glutamic Acid	GTGTAGTCGACCAGAAGGATAGGGCTCTGCAGACAGGCAGTAATACTTCTCTTTGCTCTC
Valine/Glutamic Acid	GTGTAGTCGACCAGAAGGATAGGGCTCAACAGACAGGCAGTAATACTTCTCTTTGCTCTC
Isoleucine/Glutamic Acid	GTGTAGTCGACCAGAAGGATAGGGCTCAATAGACAGGCAGTAATACTTCTCTTTGCTCTC
Glycine/Glutamate	GTGTAGTCGACCAGAAGGATAGGGTTGGCCAGACAGGCAGTAATACTTCTCTTTGCTCTC
Alanine/Glutamate	GTGTAGTCGACCAGAAGGATAGGGTTGTGCAGACAGGCAGTAATACTTCTCTTTGCTCTC
Valine/Glutamate	GTGTAGTCGACCAGAAGGATAGGGTTGAACAGACAGGCAGTAATACTTCTCTTTGCTCTC
Isoleucine/Glutamate	GTGTAGTCGACCAGAAGGATAGGGTTGAATAGACAGGCAGTAATACTTCTCTTTGCTCTC
Glycine/Asparagine	GTGTAGTCGACCAGAAGGATAGGGTTTGCCAGACAGGCAGTAATACTTCTCTTTGCTCTC
Alanine/Asparagine	GTGTAGTCGACCAGAAGGATAGGGTTTGCAGACAGGCAGTAATACTTCTCTTTGCTCTC
Valine/Asparagine	GTGTAGTCGACCAGAAGGATAGGGTTTAACAGACAGGCAGTAATACTTCTCTTTGCTCTC

Isoleucine/Asparagine	GTGTAGTCGACCAGAAGGATAGGGGTTAATAGACAGGCAGTAATACTTCTCTTTGCTCTC
Glycine/Leucine	GTGTAGTCGACCAGAAGGATAGGGAAGGCCAGACAGGCAGTAATACTTCTCTTTGCTCTC
Alanine/Leucine	GTGTAGTCGACCAGAAGGATAGGGAAGTGCAGACAGGCAGTAATACTTCTCTTTGCTCTC
Valine/Leucine	GTGTAGTCGACCAGAAGGATAGGGAAGAACAGACAGGCAGTAATACTTCTCTTTGCTCTC
Isoleucine/Leucine	GTGTAGTCGACCAGAAGGATAGGGAAGAATAGACAGGCAGTAATACTTCTCTTTGCTCTC
Glycine/Valine	GTGTAGTCGACCAGAAGGATAGGGTACGCCAGACAGGCAGTAATACTTCTCTTTGCTCTC
Alanine/Valine	GTGTAGTCGACCAGAAGGATAGGGTACTGCAGACAGGCAGTAATACTTCTCTTTGCTCTC
Valine/Valine	GTGTAGTCGACCAGAAGGATAGGGTACAACAGACAGGCAGTAATACTTCTCTTTGCTCTC
Isoleucine/Valine	GTGTAGTCGACCAGAAGGATAGGGTACAATAGACAGGCAGTAATACTTCTCTTTGCTCTC

Primer Set for YYH Mutagenesis

Mutagenic sites are underlined.

Mutation	Primer Sequence (5' => 3')
Y499G	CTGCGGGCAAGTGTAGCGCGCATAGCCCCAGGAGGACTCCATAAAGGTGTCGAAAG
Y499S	CTGCGGGCAAGTGTAGCGCGCATATGACCAGGAGGACTCCATAAAGGTGTCGAAAG
Y499R	CTGCGGGCAAGTGTAGCGCGCATAAACGCCAGGAGGACTCCATAAAGGTGTCGAAAG
Y499A	CTGCGGGCAAGTGTAGCGCGCATAAAGGCCAGGAGGACTCCATAAAGGTGTCGAAAG
Y499I	CTGCGGGCAAGTGTAGCGCGCATAGATCCAGGAGGACTCCATAAAGGTGTCGAAAG
Y499L	CTGCGGGCAAGTGTAGCGCGCATACAACCAGGAGGACTCCATAAAGGTGTCGAAAG
Y499V	CTGCGGGCAAGTGTAGCGCGCATATACCAGGAGGACTCCATAAAGGTGTCGAAAG
Y527A	CATAATGGCGTGTTCAATACCACCAATAGCGATATCCACCGGCAGCCAGTAGTTAG
Y527T	CATAATGGCGTGTTCAATACCACCAATCGTGATATCCACCGGCAGCCAGTAGTTAG
Y527G	CATAATGGCGTGTTCAATACCACCAATACCGATATCCACCGGCAGCCAGTAGTTAG
Y527L	CATAATGGCGTGTTCAATACCACCAATAAGGATATCCACCGGCAGCCAGTAGTTAG
Y527I	CATAATGGCGTGTTCAATACCACCAATTATGATATCCACCGGCAGCCAGTAGTTAG
Y527M	CATAATGGCGTGTTCAATACCACCAATCATGATATCCACCGGCAGCCAGTAGTTAG
H537G	ATTGGTGGTATTGAACACGCCATTATGGGACTGCTCTACTTCCGCTTCTTCCAC
H537T	ATTGGTGGTATTGAACACGCCATTATGACTCTGCTCTACTTCCGCTTCTTCCAC
H537F	ATTGGTGGTATTGAACACGCCATTATGTTCTGCTCTACTTCCGCTTCTTCCAC
H537Q	ATTGGTGGTATTGAACACGCCATTATGCAACTGCTCTACTTCCGCTTCTTCCAC

Data Table for MLYYH -27/SUP4 R252 Library Colonies at LRS Mutation Sites

252R Colony	Nucleotides	Amino Acids
1	ATTCTT + TCA.GCT.CAA	IL + S.A.Q
2	ATTGTA + GGC.GCT.GGA	IV + G.A.G
3	GTTGAG + GTA.GCT.ACT	VE + V.A.T
4	GCACAA + TCA.ACG.ACT	AQ + S.T.T
5	GCAAAC + GCT.ACG.GGA	AN + A.T.G
7	GGCGAG + GCT.ATG.ACT	GE + A.M.T
8	GTTGTA + CGT.ACG.ACT	VV + R.T.T
9	GCAACG + CGT.GCT.CAA	AT + R.A.Q
10	GTTAAC + GCT.GGT.ACT	VN + A.G.T
11	ATTTCC + TCA.ACG.ACT	IS + S.T.T
12	GTTCTT + GTA.GCT.GGA	VL + V.A.G
13	GCAAAC + GTA.GCT.GGA	AN + V.A.G.
14	GCACCC + CGT.GGT.CAA	AP + R.G.Q
15	GGCCAA + ATC.GCT.ACT	GQ + I.A.T
16	ATTGAG + GTA.GCT.GGA	IE + V.A.G

Data Table for MLYYH -27/SUP4 T252 Library Colonies at LRS Mutation Sites

252T Colony	Nucleotides	Amino Acids
1	GTTCTT + TCA.ATA.TTT	VL + S.I.F
2	GGCGTA + GCT.ACG.TTT	GV + A.T.F
3	ATTTCC + TCA.GCT.ACT	IS + S.A.T
4	GGCGAG + ATC.GGT.CAA	GE + I.G.Q
5	GCACAA + TCA.ACG.ACT	AQ + S.T.T
6	GTTGAG + TCA.CTT.GGA	VE + S.L.G
7	ATTCAA + TCA.GCT.ACT	IQ + S.A.T
8	G(C/T)(A/T)CTT + TTG.CTT.CAA	(A/V)L + L.L.Q
9	GCACCC + GGC.GCA.CAA	AP + G.A.Q
10	GGCCAA + GGC.ATG.ACT	GQ + G.M.T
12	GGCCCC + TCA.GGT.ACT	GP + S.G.T

LC/MS of yeGFP-wt Purified from Yeast

Mag1

C:\Users\admin\Desktop\Chatterjee lab\Alex\LC-MS15-03-16 yeGFP.bt.

NL=1.21e+006 TIC=4.31e+007 S/N=551

A 27512.0

27554.0

Mass

NL=1.05e+005 TIC=5.48e+006 S/N=167

744.584 787.069

A37 A36 A35 A34 834.710

A38 A39 A40

A41 A42 A43

A33 A32 888.495 A31 918.081 A30 949.703 A29

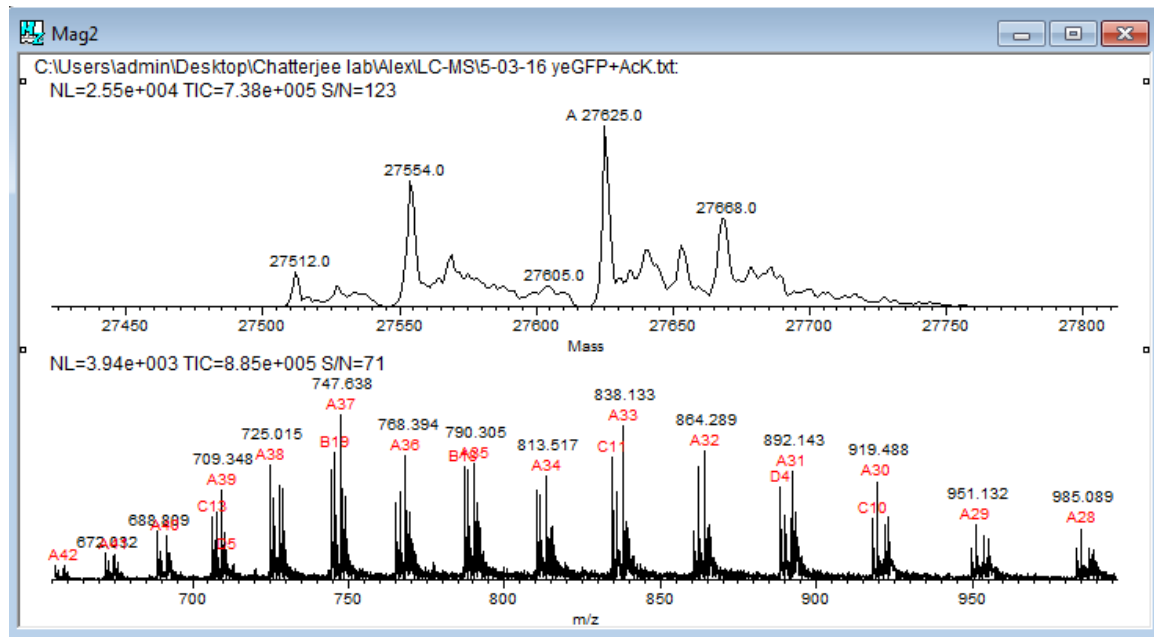
A28 1019.968 A27 A26

1101.490 1147.342 1197.173 1251.555 1311.104

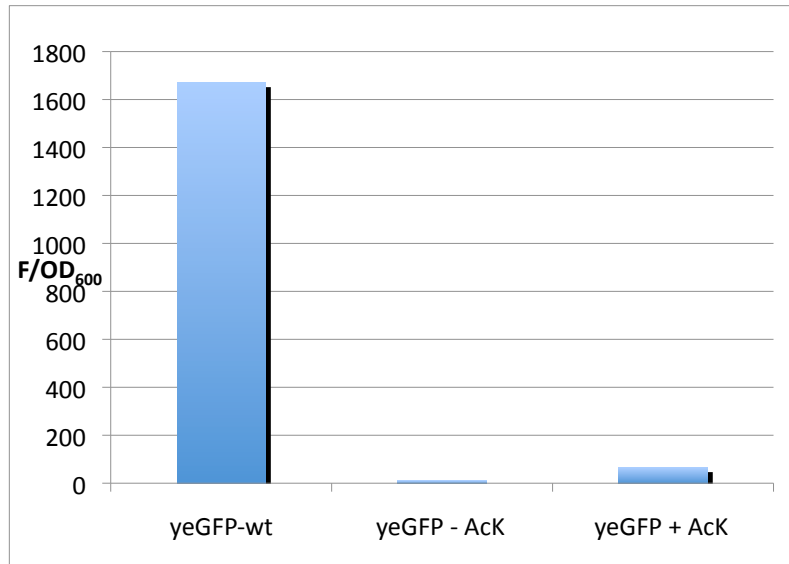
A25 A24 A23 A22 A21

m/z

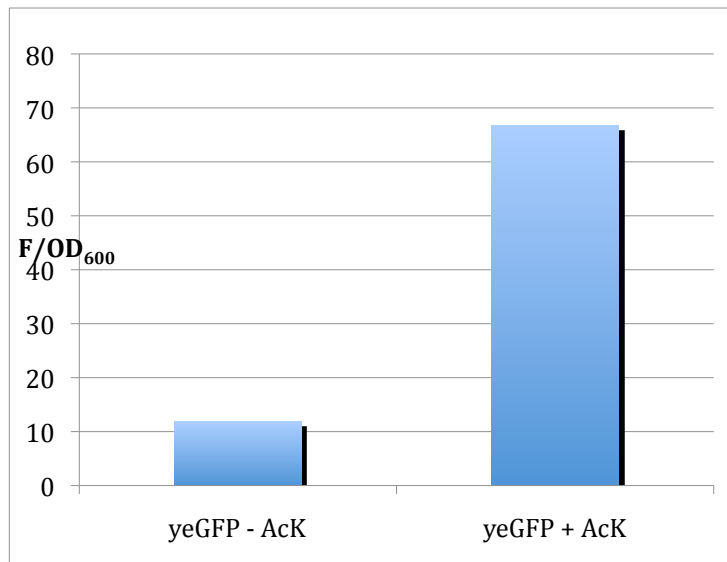
Picture includes leftover yeGFP-wt and acetylated versions of both proteins.



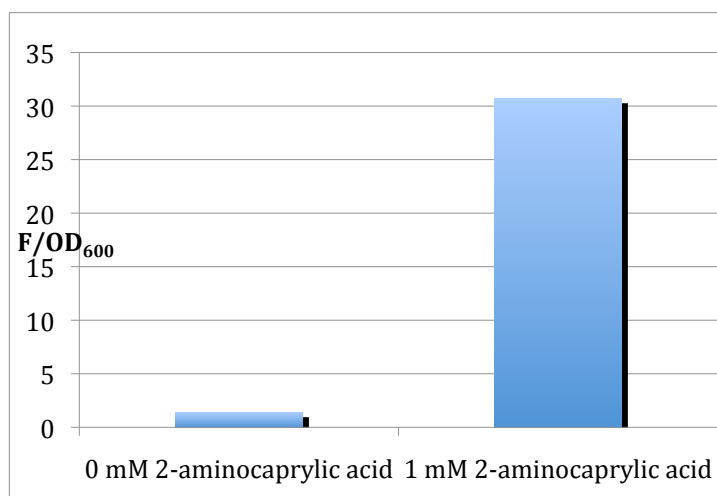
Comparison of yeGFP-wt to yeGFP-TAG Expression



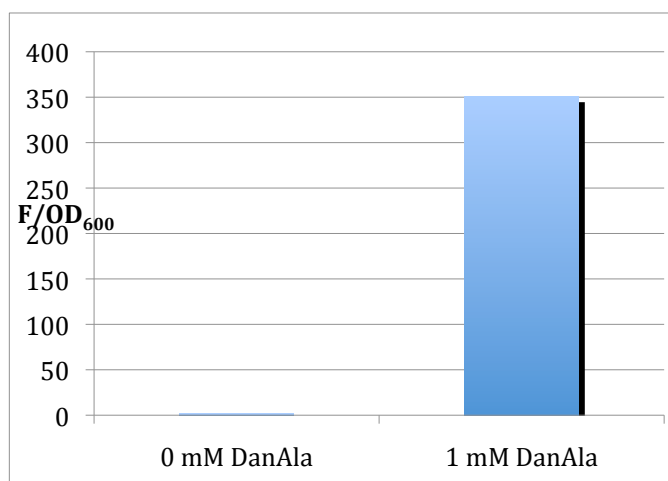
Closer Look at yeGFP - AcK and yeGFP + AcK Lanes of Previous Graph



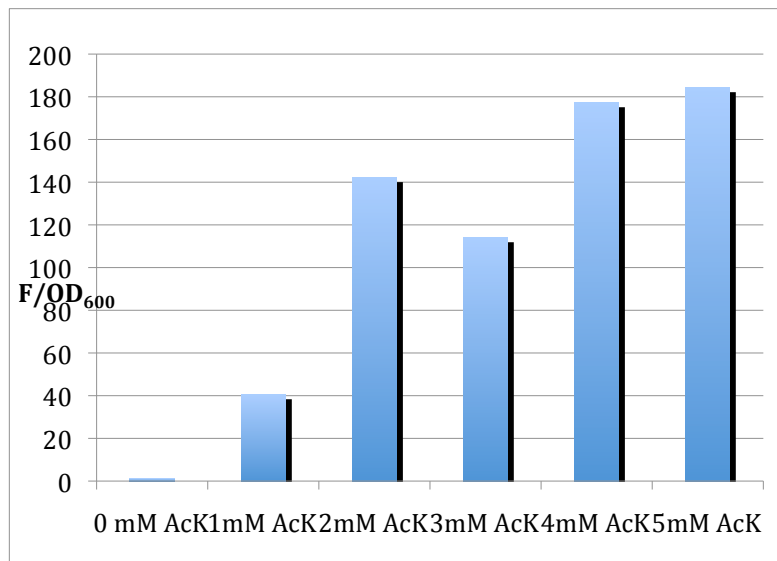
Demonstration of yeGFP + CAP Expression



Demonstration of yeGFP + DanAla Expression

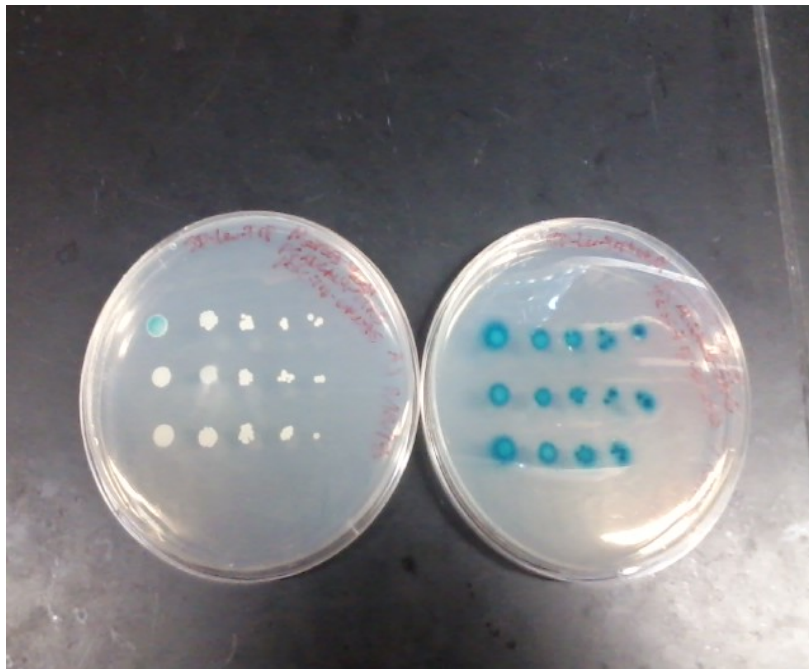


Extended Version of AcK Gradient Expression



Other Yeast Selection Tests and Spot Plates

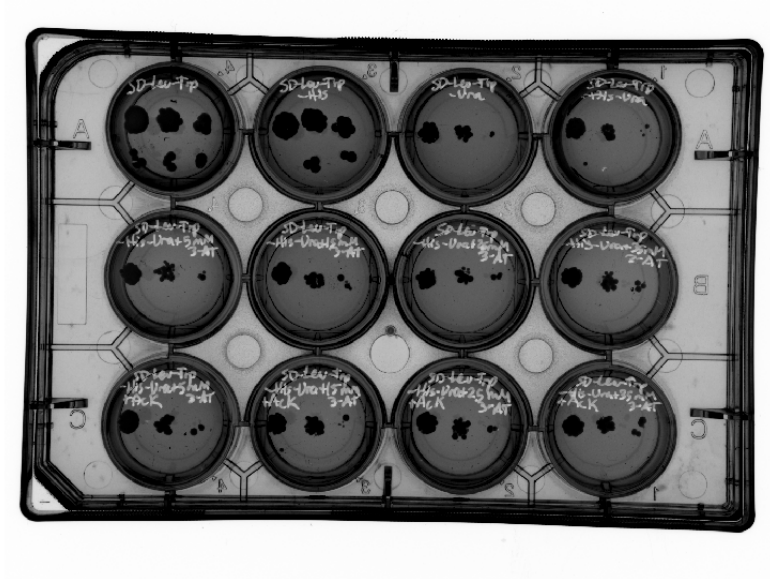
X-Gal Assay with Tyrosyl System, -OMeY (left) versus +OMeY (right) Conditions



Example of Early Double Selection Spot Plating with pESC-polyLRS-tRNA Dilution Series

The plate indicates that the yeast growth is not responsive to AcK induction and that growth largely exists on the –His and –Ura auxotrophic conditions.

2015-09-21 Double Selection Day 5



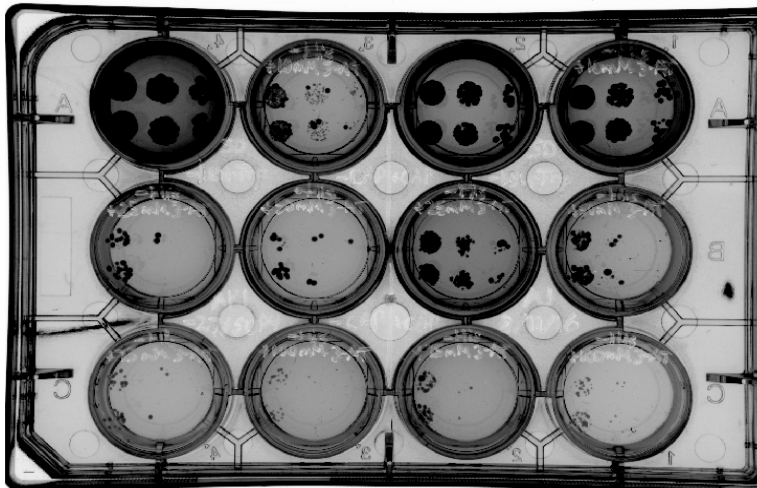
lacZ Screen with pGADGAL4-wt (Top Half) and pGADGAL4-1TAG (Bottom Half)

The left 2 wells have +X-Gal +OMeY; the middle 2 wells have +X-Gal –OMeY; the right 2 wells have –X-Gal –OMeY. pGADGAL4-wt is not a suitable positive control, but pGADGAL4-1TAG did leak and turn blue without OMeY.



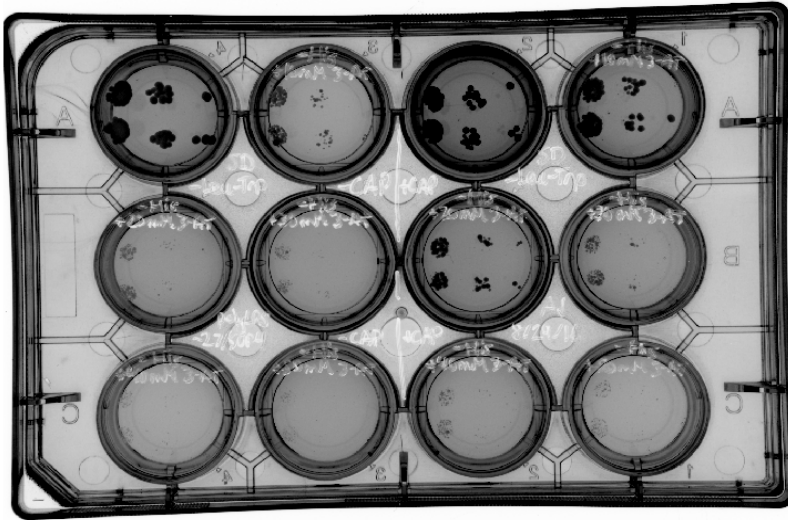
AK1 -27/SUP4 3-AT Gradient at Later Time Point

2016-08-15 AK1.27 3-AT Gradient Day 4



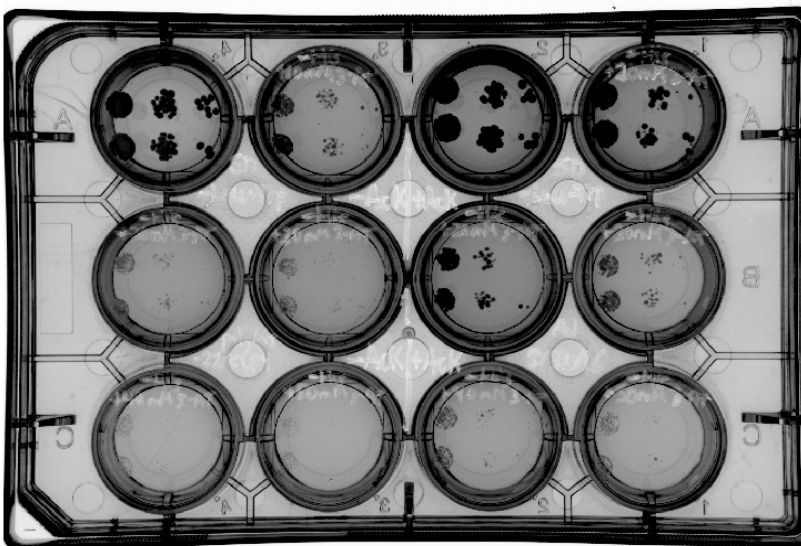
polyLRS -27/SUP4 Construct on 3-AT Gradient with CAP

2016-09-01 polyLRS 3-AT Gradient Day 3



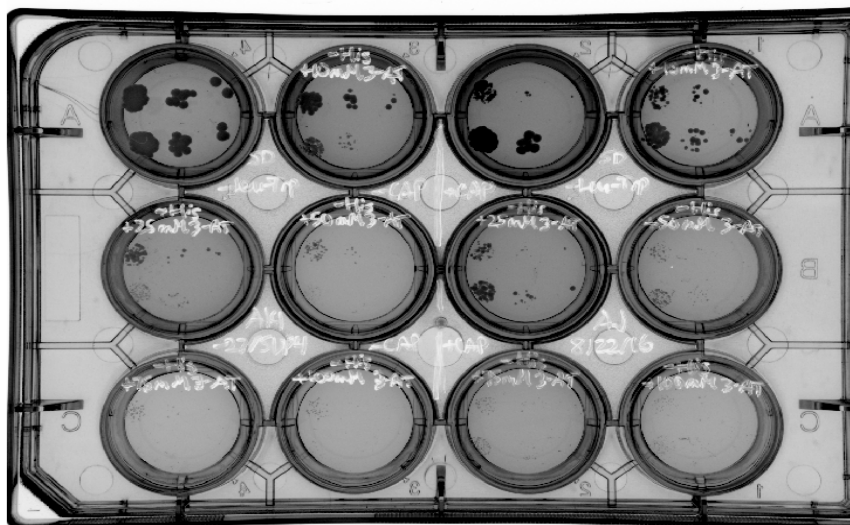
polyLRS -27/SUP4 Construct on 3-AT Gradient with AcK

2016-09-03 polyLRS 3-AT Gradient Day 3



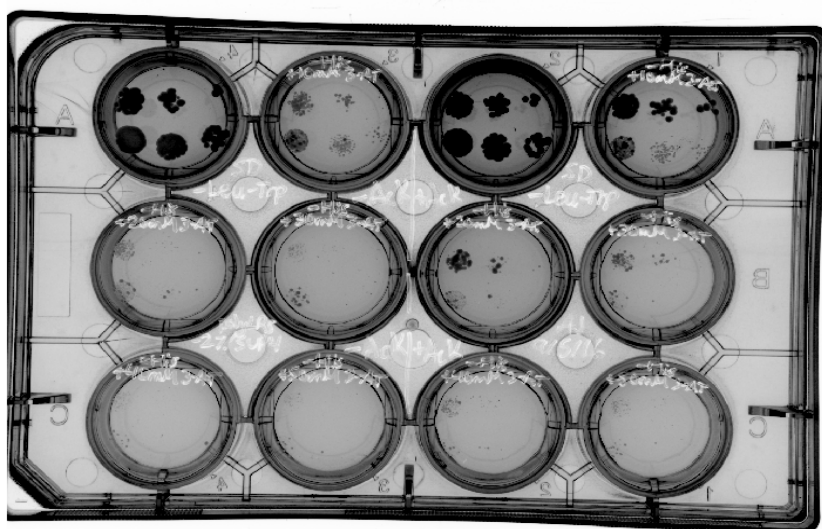
Comparison of AK1 A252 (Top Spots) and AK1 T252 (Bottom Spots) on 3-AT Gradient

2016-08-25 27.SUP4 252T Comparison Day 3



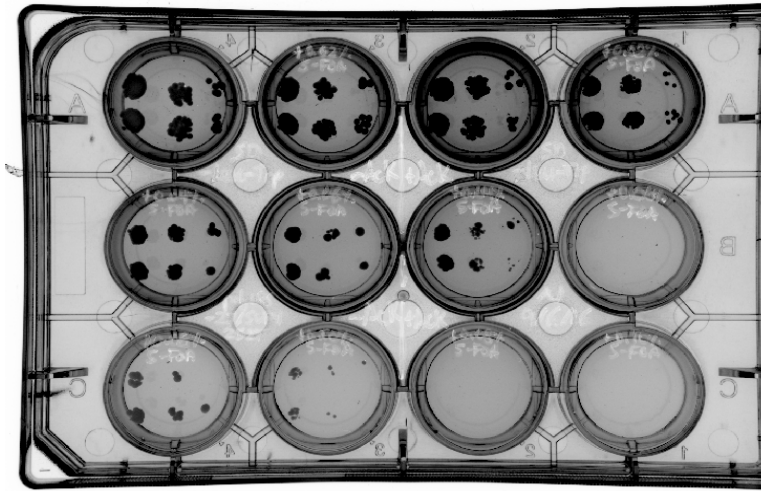
Comparison of polyLRS R252 (Top Spots) and polyLRS T252 (Bottom Spots) on 3-AT Gradient

2016-09-08 polyLRS 252T Comparison Day 3

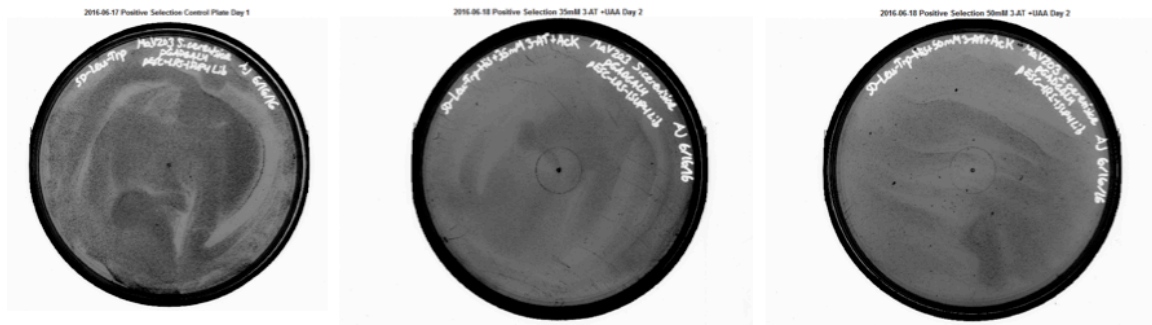


polyLRS -27/SUP4 5-FOA Gradient

2016-09-09 polyLRS 5-FOA Gradient Day 3

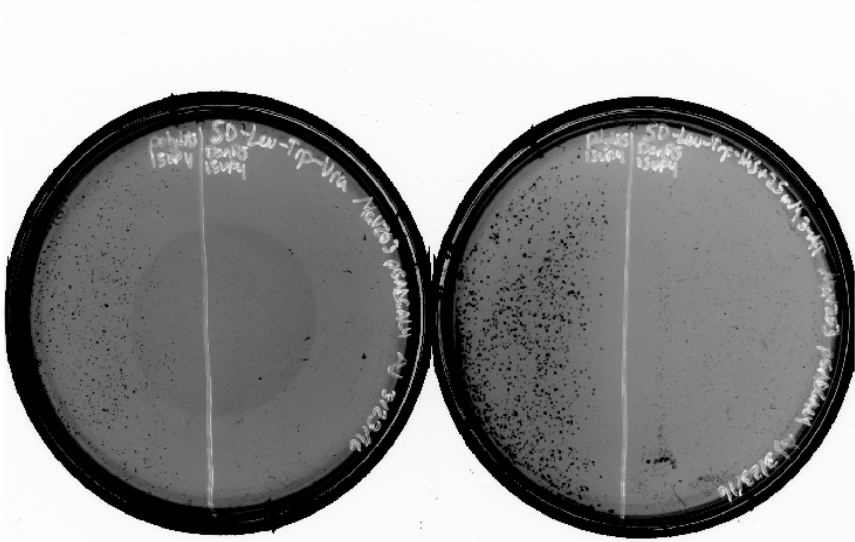


Example of MLYYH LRS 1SUP4 Library Positive Selection



Comparison of polyLRS and DanRS, with 1SUP4 Constructs on Two Positive Selections

2016-03-26 Positive Plates (-UAA)(poly vs Dan) Day 3



2016-03-26 Positive Plates (+UAA)(poly vs Dan) Day 3

